Readers' attitudes to self-archiving in the UK

Rhiannon Macfie Miller

MSc Dissertation

October 2006

School of Communication Arts, Napier University

Abstract

The online self-archiving by authors of their scholarly articles has been proposed as an alternative to author-pays open access publication, but has the potential to undermine journal publisher income if the ready availability of self-archived articles leads to a drop in subscriptions. This study investigated the awareness of self-archiving and use of self-archived articles in a survey of a mainly academic population including both authors and non-authors, and looks at their attitudes to self-archived papers and whether they view them as an authoritative alternative to subscription access. In total, 70% of respondents had heard of self-archiving, though only 15% knew a lot about it, and 71% had used self-archived papers. These proportions are higher than in previous studies, suggesting that awareness has grown. Most self-archived papers used came from websites rather than repositories, particularly among those whose awareness of self-archiving was low. Use of self-archived articles was greater amongst those who had published more papers and also depended strongly on subject field – use and awareness were both particularly low in the field of medicine. People who were more aware of self-archiving were less likely to view the publisher's official version as the only authoritative version and more likely not to care about the online location of articles. Moreover, authors who had self-archived tended to archive the publisher's official version regardless of whether they were permitted to. These results suggest that the awareness of self-archiving is currently mostly limited to academic authors and is unlikely to grow beyond this in the short term. However, in the long term, the combination of high rates of self-archiving of the publisher's official version, coupled with the devaluation of the journal as the authoritative source of material together with increased convenience of access to self-archived material, could result in fewer people accessing articles through subscription-based methods.

Acknowledgements

I would like to thank the following people for their help: my husband Robin Miller, for suggestions, support, and cooking and cleaning single-handedly for several weeks; Patrick Miller, for advice on data analysis and statistics; my mother, Dr Elizabeth Evans, for help in sourcing relevant literature and provision of email addresses for the survey; and David Williams, Ben Williams, and Robin Miller (again) for help with the collection of email addresses.

Contents

Abstract ii
Acknowledgements iii
Contents iv
List of figures and tables v
Introduction 1
Definitions 1
Literature review 3
Search methods 3
Definition of Open Access 6
'Gold' open access: free journals 6
'Green' open access: self-archiving 9
Current status 14
Questions remaining 18
Methods 20
Rationale 20
Aims and hypotheses 20
Questionnaire design 21
Sample selection 23
Results and discussion 24
Sample demographics 25
Searching behaviour 35
Knowledge and use of self-archiving 43
Author behaviour 70
Comments 84
Conclusions 86
Reader use of self-archived articles 86
Author self-archiving behaviour 87
Perception of repository value 87
Implications for scholarly publishing 88
Limitations of the study 89
Further research 89
References 90
Appendix 1: Web survey 102
Invitation email 102
Survey text 103
Appendix 2: Analysis of institutional repositories 113
Introduction 113
Survey of institutional repositories 113

List of figures and tables

Table 1: Institutions of respondents 25
Table 2 Institutional repository status 26
Figure 1: Numbers in the institutions (aggregated version) 27
Figure 2: Institutional repository status 28
Figure 3: Numbers in the various subject fields 29
Figure 4: Institutional repository presence by subject field 30
Figure 5: Numbers in the various roles 31
Figure 6: Number of years in field 32
Figure 7: Relationship between role and time spent in field 32
Figure 8: Subject area and years spent working in it 33
Figure 9: Number of papers used per year 34
Figure 10: Relationship between subject area and number of papers used per year 34
Figure 11: Number of papers published by authors in past year 35
Table 3: Search engine and database properties 37
Table 4: Choice of search engines 38
Figure 12: Overall properties of search facilities used 39
Table 5: Preferred search facility 40
Figure 13: Preferred search facility properties 41
Figure 14: Preferred search facility properties ('Internal' not defined) 41
Figure 15: Choice of online paper formats 42
Figure 16: Preferred paper format 43
Figure 17: Knowledge of self-archiving 44
Figure 18: Awareness of self-archiving in different subject fields 45
Figure 19: Effect of number of papers used on awareness of self-archiving 45
Figure 20: Awareness of self-archiving amongst respondents who had and had not published a paper 46
Figure 21: Awareness of self-archiving among authors by papers written in the past year 47
Figure 22: Understanding and valuing of repositories 47
Figure 23: Understanding of repositories affected by knowledge of self-archiving 48
Figure 24: Use of self-archived material 49
Figure 25: Knowledge of self-archiving and use of self-archived material 50
Figure 26: Effect of subject field on use of self-archived articles 51
Figure 27: Use of self-archived material by role of respondents 52
Figure 28: Number of years in field and use of self-archived material 52
Figure 29: Use of self-archived material by number of papers produced in previous year 53
Figure 30: Search facilities and use of self-archived material 54
Figure 31: Knowledge of self-archiving and choice of search facility 54
Figure 32: Use of self-archived material among users of the most popular search facilities (overall average = 71%) 55
Figure 33: Number of respondents using multiple search facilities 56
Figure 34: Perceived repository status 56
Figure 35: Knowledge of institutional repositories by knowledge of self-archiving 57
Figure 36: Perceived and actual institutional repositories 58
Figure 37: Use of articles from institutional repositories 59
Figure 38: Knowledge of self-archiving affects use of institutional repositories 60
Figure 39: Use of institutional repositories and perception of institutional repository status 61
Figure 40: Awareness and use of subject-based repositories 61
Figure 41: Use of subject-based repositories by subject 62

Figure 42: Use of articles from repositories and authors' websites 63
Figure 43: Awareness of self-archiving affects accessing from non-publisher sources 65
Figure 44: Subject influence on use of articles from non-publisher sources 66
Figure 45: Attitudes to factors relevant to self-archiving 67
Figure 46: Effect of awareness of self-archiving on views of self-archiving 69
Figure 47: Publication year of last paper 71
Figure 48: Copyright policies of publishers 72
Figure 49: Self-archiving policies of publishers 72
Figure 50: Time before an article can be self-archived 73
Figure 51: Open Access policies of publishers 74
Figure 52: Author perception of permissions 75
Figure 53: Authors' knowledge of self-archiving and perception of self-archiving permissions 75
Figure 54: Accuracy of authors' perceptions of self-archiving permissions 76
Figure 55: Effect of knowledge of self-archiving on accuracy of authors' perception of self-archiving policy 77
Figure 56: Accuracy of authors' perception of copyright ownership 77
Figure 57: Accuracy of authors' perception of embargo policies 78
Figure 58: Methods of self-archiving of most recent paper 79
Figure 59: Effect of awareness of self-archiving on intention to self-archive current paper 79
Figure 60: Attitudes and action of self-archiving 80
Figure 61: Reasons against self-archiving 81
Figure 62: Reasons for self-archiving 82
Figure 63: Difference between archived and journal version of last self-archived article 82
Figure 64: Preferred format for uploading self-archived papers 84
Table 6: Repository status of institutions 115
Table 7: Properties of institutional repositories 117
Table 8: Publicity of repositories 118

Introduction

In the world of academic and scholarly publishing, particularly in the fields of science, technology, and medicine, a debate has for some time been under way about the best way to publish research articles. Campaigners argue that the results of original research should be available for all to view and publishers should not charge for access to this material, especially since they do not pay the authors for it. Publishers argue that they fulfil a vital function in managing the peer review of these articles and ensuring that only high-quality research is published, and this needs to be paid for.

Several prominent research funders, including the National Institutes of Health, the Wellcome Trust, and Research Councils UK, are now demanding that all articles by researchers they fund be freely available. Some journals now offer a service whereby individual articles can be made available to all on payment of a fee by the author (or the author's institution or sponsor), but many authors are not willing to pay this. Many authors make copies of their articles available from their personal website, and in a formal extension of this, in some subject fields there are centralised archives where these can be published and downloaded. More recently, institutions, particularly universities, have begun to set up their own archives. The advantages of these archives from the point of view of the researcher are that they make all the data about the articles available, they are easily (and collectively) searchable using common search facilities such as Google, and that they make their content freely available.

Journal publishers, especially small society publishers, fear that if people begin to use these facilities, they will stop subscribing to their journals, and income will fall. This dissertation looks at whether, and what sort of, people are using these facilities, and, if they are, for what reasons. This will give a better idea of whether use of these facilities is indeed likely to affect journal publishers' income.

Definitions

In this dissertation, the following definitions will be used:

Open access Any form of publication which means that articles in a journal can be accessed for free by people who do not subscribe to that journal.

Open access journal A journal which makes all its articles available free to non-subscribers as a matter of policy.

Self-archiving The publishing by an author of his or her own work, on a personal or institutional website or in a repository, in addition to its publication in a journal.

Repository An archive of publications in a certain field (subject repository) or at a certain institution (institutional repository), with searchable and freely available data including full-text articles for download.

Author-pays A system of publishing whereby the author (or the author's institution or sponsor) pays a fee to the journal publisher in order that the publisher make the article available online to non-subscribers as well as subscribers.

Preprint A pre-publication draft of a paper, not including changes suggested by peer reviewers.

Postprint The final draft of a paper, including changes suggested by peer reviewers but not formatted or copy-edited by the journal publishers.

Journal PDF The official publication version of a paper, including peer-review changes and copy-editing, formatted to comply with the journal style and made available from the journal publishers.

Literature review

Search methods

I conducted full-text searches in ProQuest, Emerald, Ingenta, ScienceDirect, Metapress and Google Scholar, with default sorting, using the following keywords:

- open access
- "open access"
- open access publishing
- "open access" publishing

In addition to this, I also sourced literature by the following methods:

- I browsed the print abstracts publication *Press* for relevant articles. This highlighted that the journal *Learned Publishing* was a particularly rich source of material, and so I searched in this journal for articles and continued to monitor further issues of the journal.
- I took relevant articles from the results of a separate search conducted by Lucy Harrier in March 2005. These articles mostly came from open access bibliographies available on the E-prints archive (http://openaccess.eprints.org/).
- I monitored the *Open Access News* weblog (http://www.earlham.edu/~peters/fos/fosblog.html), and browsed the archives for further relevant information. This pointed me to the subject repository *E-LIS* (E-prints in Library and Information Science; http://eprints.rclis.org/), which was a useful source of material, particularly book chapters and theses.
- I monitored general news sources such as the BBC and *The Economist* and followed up reports relating to open access.
- The BMJ was monitored by Dr Elizabeth Evans, who forwarded me relevant articles published there.
- Where studies and policy statements were discovered only through news reports, I attempted to track down the original. If a full reference was not given in the report, I searched in Google or Google Scholar on the study title, authors, and/or institutions as reported in the news article in order to obtain the original article.
- I followed up references cited in some articles to find further relevant studies or announcements.

- I continued to search Google and Google Scholar regularly through the course of writing this dissertation for further information on topics raised in the articles.

I narrowed down these articles using the following criteria:

- News items reporting the results of a study were discarded if that study was also available and no new information was added. This also applied to news articles reporting policy statement announcements and press releases.

- Editorials announcing the implementation of open access policies for an individual journal, or explaining why the journal was not intending to implement open access policies, were discarded.

- Overviews acting as simple introductions to the concept of open access were for the most part discarded.

- Duplicate articles that appeared in more than one resource were merged.

In total, 179 unique references were obtained.

The table below summarises some of the articles most relevant for this dissertation.

Authors	Article type	Summary of key findings
Lawrence, 2001	Study	Open access publication can increase the impact of a paper by up to 300%.
Willinsky, 2003	Speculation	Loss of revenue is not necessarily a bar to open-access publishing for learned societies as they can make cost savings and gain revenue through advertising.
Swan and Brown, 2005	Survey	Self-archiving is not commonplace amongst authors. Authors fear that publishing their articles in open access form could affect their funding or careers.
Anderson, 2004	Speculation	Authors may not wish to publish in open access form because they are not willing to pay charges or because open access journals are seen as being of lower quality.
Harnad and Brody, 2004	Study	Open access articles (self-archived) are cited more than non-open access articles in the same journals.
Harnad *et al.*, 2004	Commentary	90% of journals allow self-archiving but only 20% of authors do this.
Lamb, 2004	Speculation	Market forces will encourage journals to move to open access models
Goodman, 2005	Speculation	6 cases examined based on legislative scenarios and consequences of policy decisions extrapolated. If self-archiving becomes mandatory, journals may survive if authors still prefer to access the journal version of an article.
Nicholas and Rowlands, 2005	Survey	Authors see reputation of the journal and impact factor as most important criteria in deciding where to publish. Permission to post article elsewhere is low on the list. Many authors are unwilling to place their articles in repositories.
Gannon, 2005	Speculation	Open access journals may cause financial difficulties for learned societies publishing competing journals.
Ware, 2006	Survey	The entire content of a journal needs to be available in repositories before librarians will consider unsubscribing for this reason.

Kaufmann-Wills Group, 2005	Study/Survey	Detailed study of current state of affairs. 40% of open access journals run at a loss. Many traditional journals see open access as a threat. Peer review with open access may be less rigourous.
Kurtz *et al.*, 2005	Study	Self-archived articles are more frequently cited not because they are freely available but because they are available sooner and because authors promote their most important (most citeable) articles by selectively publishing them in open access form.
RCUK, 2006	Statement	Requires authors funded by RCUK to publish in open access form.
Rowlands and Nicholas, 2005b	Survey	29% of senior authors have published in an open access journal. Very few authors are willing to pay personally or to have much of the cost taken from their research funding.
Schroter *et al.*, 2005	Survey	Authors are not keen on author-pays but would be willing to submit to high-quality journals even if they charged for publication.
Suber, 2005	Speculation	Speculation on the motives of self-archiving authors in response to Wren (2005).
Wren, 2005	Study	A search for PDFs freely available online (other than in the original journal) found that more of them came from higher-impact journals.

Definition of Open Access

Open access is defined by the Budapest Open Access Initiative as a system of publishing by which users may

> read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.
>
> (Open Society Institute, 2001)

This means that in making an article available through open access, the publisher not only does not charge the reader to access the article but also does not restrict the copying or redistribution of the article in any way.

Open access has been steadily becoming more widespread over the past few years (a good comprehensive history of the beginnings of open access can be found in Poynder, (2004a)). While the demand for open access publication has been fuelled by the rising subscription costs of traditional journals, advocates of the system argue on principle that research will proceed faster and more efficiently if researchers can maximise their access to the literature (Velterop, 2004). There is some literature indicating that articles published through an open access method have more citations than non-open access articles (Lawrence, 2001; Harnad and Brody, 2004). Funding bodies such as the Wellcome Trust have begun to mandate that research funded by them be published under an open access model, and governments of various countries have debated regulating STM research publishing with a view to making the results of research more accessible.

Open access publication is generally seen as a threat by traditional publishers, whose income is derived from the subscription fees paid for access to the journals. However, many researchers and research funders believe that it is important to make access to the results of research as easy as possible in order to further future research.

The two main variants of open access commonly recognised are the 'gold model' in which an article is published by a journal which makes the article freely accessible online; and the 'green model' whereby the article, having been published in a journal under the traditional publishing process, is then also made available elsewhere (Guédon, 2004).

'Gold' open access: free journals

Under the 'gold' form of publication, a journal makes its articles available to everyone, rather than limiting access to subscribers only. This has become known as 'author-pays' publication, from what is seen as one of the main features of this form of open access: in order to cover the costs of publication, the author pays a fee to the journal publishers to have the article published. This may be a fully open access journal where all articles are freely available to all potential readers, or it could be a traditional subscriber journal which offers the option of charging at the point of publication for a specific article to be

made available, while the rest of the articles are published in the traditional manner.

'Author-pays' is, however, somewhat of a misnomer. It is not generally the author who pays but the agency that has funded the work or the institution in which the author is based (Harris, 2005). In addition, a large proportion of open access journals do not charge the author a fee at all, but source their income from advertising or from public sponsorship; others work with a mixture of income streams (Kaufmann-Wills Group, 2005). In a survey of authors, only 26% associated the term 'author pays' with open access journals (Rowlands and Nicholas, 2005b), and only about the same proportion (29%) would be prepared to pay for such publication personally.

Over the past few years, a number of open access journals have started up, including the high-profile launch of the Public Library of Science journals such as PLoS Medicine and PLoS Biology. Some other traditional journals have moved to an open access service, particularly society-owned journals, and major journal publishers such as Reed Elsevier, Blackwell Publishing and Springer Verlag now offer an optional author-pays service in some of their journals.

Open access journals

The Directory of Open Access Journals (DOAJ; http://www.doaj.org) in January 2006 listed 2,007 peer-reviewed open access journals. The Kaufmann-Wills report (Kaufmann-Wills Group, 2005) studied the properties of open access journals listed in the DOAJ in some depth, and compared them with journals using other access models taken from the lists of journals belonging to the sponsoring societies, the Association of Learned, Professional and Scholarly Publishers and the American Association for the Advancement of Science. One of its main findings was that only 35% of open access journals are turning a profit, while 25% are breaking even and 40% making a loss. While it is reasonable for a journal to fail to make a profit for its first few years as starting costs are overcome (Morris, 2005), and a majority of open access journals are very new starts indeed, the report also concludes that there is little likelihood of the 65% currently not running at a profit improving their financial status in the foreseeable future.

The same report also found that open access journals, while still peer-reviewed, were more likely to be less rigorous in their reviewing, with many carrying out the review in-house rather than using the traditional system of external peer-review. They were also less likely to carry out copy-editing functions. This may partly be because open access journals tend to be published by smaller organisations – individuals, non-profit groups or academic departments rather than commercial publishers or learned societies, who would have fewer resources to spare; and also because, again, many of them are new starts who have so far failed to gather sufficient fame to be able to draw on a pool of willing external volunteer referees.

Rowlands *et al.* (2004), who conducted a major survey of published authors, suggests that 74% of senior authors would not choose to publish in an open access journal. Harris (2005) criticises this study (in its preprint incarnation) on the grounds that while the study defined 'open access journals' as journals to which a fee has to be paid to have a paper published, this was not made clear to the survey respondents. He also points out that in the survey 34% of respondents claimed to know 'nothing at all'

about open access.

Harris suggests that many authors assume that 'open access journals' are by definition poorly peer-reviewed or not peer-reviewed at all, or that they personally (rather than their funding agency or institution) would have to pay the publishing fee. Because the authors were commenting from a position of ignorance, Harris concludes, the survey results are, if not invalid, then at least not as clear-cut as they might appear. However, the same survey (Rowlands *et al*., 2004) found that authors tended to associate open access publishing with high quality and not with authors paying to publish, which would appear to negate Harris' claims.

Nevertheless, Harris does have a point. The question regarding what properties respondents associated with open access came before the question on whether they were prepared to publish in an open access journal. The question included a definition of open access containing the phrase 'The costs are met by charging *authors*' [emphasis mine] and was couched in terms not of 'would you be willing to publish' but 'what would you consider a reasonable payment to have your paper published' in an open access journal; one of the response options was '*I'm* not prepared to pay anything' [emphasis mine]. The wording of the question would clearly imply to many respondents that they, personally, would be expected to pay in this hypothetical scenario though the authors point out elsewhere (Nicholas and Rowlands, 2005) that the comments made by the respondents did not seem to indicate that they had interpreted the question in this way. It is notable, however, that in a similar survey conducted by the same authors the following year (Rowlands and Nicholas, 2005a), the definition of open access had been changed to 'Open access publishers sometimes meet their costs by charging authors *(usually through the author's funding body or employer)*' [emphasis mine].

In this second survey, authors were asked only whether they had published in open access journals, not whether they were willing to. Twenty-nine per cent said that they had, but of these, less than half chose to publish in an open access journal on principle. For most authors, the most important factor in deciding where to publish is the prestige of the journal or its suitability in terms of whether the paper will be read by other researchers in the same field, rather than its access model. However, this works both ways; many authors would still submit their paper to a high-quality journal even if it charged for publication Schroter *et al*., 2005.

The difference between the author attitudes shown in Rowlands and Nicholas, 2005a and those shown in Rowlands *et al*., 2004 deserves further comment. It is worth bearing in mind that even over the course of a single year the knowledge claimed by authors of open access publishing increased: while in the 2004 study 34% of authors claimed to know 'nothing at all' about open access, in the 2005 study this had decreased to 19%. This increase in knowledge would surely result in changed attitudes. Certainly a much larger proportion of authors in the 2005 study had actually published a paper in an open access journal (29% in 2005 compared with 11% in 2004). The two surveys were conducted using random samples from the same database (with those who had taken part in the first survey excluded from the second), so it is probable that the two sets of respondents are indeed comparable. However, as noted above, the questions asked were different; even where the questions were the same, the responses may have been coloured by previous questions. For example, in the 2005 survey the respondents were asked a series of questions about the importance of peer review before they were asked about open

access, and this may have meant that their responses on open access were affected by their view of how peer review is treated in open access journals more than would otherwise have been the case.

In a separate survey, authors were asked who they thought *should* pay for publication (Rowlands and Nicholas, 2005b). The majority felt that authors should make no contribution, with research funders, institutions and central government favourites for covering a large portion of the costs. Interestingly, few authors felt that a single body should pay all the costs, and many felt that readers should make a 'small contribution'.

'Green' open access: self-archiving

The process of making copies of an article available to colleagues who do not have subscription access to the journal in which it is published predates the use of the internet to share documents; generally an author was given a limited number of offprints of their paper to keep or give away to interested readers (who had to contact the author personally in order to obtain a copy). The difference between this and authors making electronic copies available through the internet ('green' open access) is that the latter allows for unlimited copying and much easier access.

Many authors who make their papers available in this way do nothing more than upload the PDF of their paper to their own personal website (Swan and Brown, 2004a), but the Open Archives Initiative and similar organisations aim for a more formal approach in which the articles are uploaded to an institutional or subject-based respository (Harnad, 2003). These repositories would ideally have compatible formats that would enable multiple repositories to be queried simultaneously with one search engine query.

The Eprint archive registry (http://archives.eprints.org/) in January 2006 showed 607 open access archives which meet these criteria, and this number is growing rapidly (Carr and Harnad, 2005); by the end of December 2005 more than 700,000 articles had been registered (though not all of these are necessarily full-text versions). It is notable that some subjects attract more self-archived articles than others. This is partly for historical reasons, as some of the oldest repositories are subject-based, for example ArXiv, the physics repository, which started life as a forum for sharing preprints in a fast-moving field where it was important to get the results published as fast as possible, and verified through peer-review later Montonen, 2005. While many archives now host both pre-prints and post-prints, some authors feel that this sharing of non-peer-reviewed work has led to a degradation in quality (Rowlands and Nicholas, 2005b), and think that there could be confusion between different versions of an article Nicholas and Rowlands, 2005.

Chemistry is an outlier among the sciences in having few open access journals (McVeigh, 2004) and a smaller number of self-archiving authors compared with physics and the life sciences (Rowlands and Nicholas, 2005a). Awareness of open access, self-archiving and the existence of repositories also differs in different fields of work, with authors in physics, astronomy, computing, business and economics most likely to have made a paper available on the Web (Rowlands *et al.*, 2004). These are fast-moving fields where authors produce large amounts of data. Researchers in the field of computing are also more likely to do more of their communication online and feel more comfortable about self-

archiving on the internet, and are more likely to be aware of open access issues in general. Medicine and the biological sciences are fast-growing fields in open access terms as well, fuelled by public demand for research on health-related issues to be made freely available.

Motivations for self-archiving

Despite the fact that the concept of self-archiving as the basis for open access has been around for over a decade (Poynder, 2004a), only a small proportion of research articles are currently self-archived (Harnad *et al*., 2004; Swan and Brown, 2004a), and Rowlands and Nicholas, 2005b show 38% of authors as being unwilling to do so. However, Swan and Brown, 2005 suggest that as many as 49% of authors may have used some form of self-archiving for at least one article during the period 2002–2005. The remainder may choose not to because they perceive it to be difficult or time-consuming, because they are unaware of any suitable repositories, or because they believe that the terms and conditions of publication forbid it, but according to Swan and Brown, 2005 the majority are simply unaware that it is an option. Authors' knowledge of self-archiving is certainly more limited than their awareness of open access journals (Rowlands and Nicholas, 2005b).

Few are unhappy with the idea of self-archiving in principle: 81% would willingly do so if required by their institution or funding body (Swan and Brown, 2005), and less then 10% would seek alternative funding rather than comply (Swan and Brown, 2004b). Nevertheless, it is worth noting that in a department which does mandate the archiving of all articles in a local repository, only 50% of papers published from that department had been deposited there by 2004 (Carr and Harnad, 2005). In addition, an older survey by Swan and Brown, 2004a indicates that authors fear that self-archiving might adversely affect their funding or careers (possibly through the confusion caused by having different versions of the same paper extant).

Wren, 2005 notes that papers from journals with a high impact factor are more likely to be made available online. There is no apparent correlation in that study between availability and copyright ownership or permissions; in other words, authors may be self-archiving their work in ignorance or defiance of the conditions imposed by the publisher. However, Wren did not differentiate in his study between author self-archiving and third-party publishing, so it is possible that readers may have made the papers available either for their own use or that of colleagues (Suber, 2005). If so, this would also be contrary to copyright restrictions for many papers.

Wren also did not differentiate between papers in an institutional or subject-based archive and papers published informally on websites. Swan and Brown, 2005 note that of authors who have used some form of self-archiving, more chose to place their article on a personal or institutional website than opted for the more formal repository archiving, and Gadd *et al*., 2003b found that 81% of researchers who accessed freely available papers did so via individual websites. However, the number and availability of repositories has increased in recent years, so it is possible that a higher proportion of recent articles are now being placed in repositories, and that the apparent higher use of personal websites is a historical artefact related to the fact that older papers were published before repositories were widely available or known about.

Wren found that authors self-archived their articles even when they were already available as open access articles from the publisher. In his commentary on Wren's study, Suber, 2005 offers as an explanation the possibility that a major motivation for self-archiving is the desire of an author for self-promotion: making their success known by making their paper more widely available; this also explains why it is papers from high-impact journals that are preferentially self-archived. This is confirmed by Kurtz *et al.*, 2005, who also found that authors tend to preferentially publish their most citeable papers in open access journals.

Open access publishing through the self-archiving method may be more attractive to authors than the gold road of open access journals for several reasons. In addition to the fact that authors do not have to pay or set aside part of their research grant to make their article freely available by this method, they can also choose from a much wider range of journals than might otherwise be available to them if they were limited to journals which offered open access themselves. This means that the author is free to choose the most relevant or most prestigious journal that will accept their paper, meaning that it has a higher importance to the reader. They still retain the benefits of the higher citation rate of open access papers (Harnad and Brody, 2004).

In addition, self-archiving could theoretically also be more attractive to publishers, as they retain their existing subscription-based funding models (Hunter, 2005). However, if the repositories become the main means of access to research articles, subscription revenues may drop. The repositories would thus be in direct competition with the published journals. Rowlands and Nicholas, 2005b asked authors whether they thought that a repository-based system would be likely to undermine the current system, and whether they thought this would be a good thing. Of those who expressed a definite opinion, twice as many (46%) thought it would be both likely and a bad thing as chose any other set of options.

The initial reaction of major journal publishers to the prospect of self-archiving was to develop their own proprietary online platforms where articles are made available to subscribers only, such as Elsevier's ScienceDirect and Blackwell Synergy. This was intended to discourage researchers from looking elsewhere for self-archived articles (Poynder, 2004b), as they would more conveniently be able to search the publishers' own site, and access the articles through their institution's subscription. Poynder suggests that this was the main driving force behind the 'Big Deal' subscription policy, as in order for this strategy to work the institution's subscription licence would have to cover a large number of journals, so that researchers would feel they had everything they needed without having to look elsewhere. This policy has, however, backfired to some extent, as the resultant increased subscription prices have fuelled the demand for open access.

At the same time, the terms and conditions publishers imposed upon authors grew steadily more restrictive (Gadd *et al.*, 2003c), with many publishers forbidding authors to, for example, post the PDF of their article on a website. Because of recent changes in policy by academic funders (RCUK, 2006; Wellcome Trust, 2005), however, many publishers have altered their terms and conditions to authorise

at least some form of self-archiving[1], although in many cases the embargo period (time after publication during which the article may not be published elsewhere) is the maximum permissible by the funding bodies (generally one year). Many self-archiving authors disregard these terms and conditions in any event; where they are unsure whether they have permission, they go ahead anyway rather than seek clarification (Swan and Brown, 2005).

Motivations for readers

The question then remains: will self-archiving actually make any difference to the way researchers work? Institutional or subject-based repositories are only likely to be a threat to the economics of traditional publishing if researchers turn to them for information in place of subscribing to journals (Goodman, 2005). While most subscription income for journal publishers comes from libraries, if libraries can provide their readers with substantially the same service without having to pay subscription fees, this is likely to prove very attractive to them, even if some of their funding now goes to the maintenance of an institutional repository.

Waltham, 2005 claims that 'if author posting of articles on open archive repositories means that, for example, a search with Google Scholar can find the free version of any or most of the articles in a journal, then over time subscription cancellations are inevitable.' This is only likely to happen, however, if readers feel that the self-archived version is at least as trustworthy as the version published in the journal, and if readers use search engines such as Google Scholar which include the free versions, rather than proprietary search engines run by the publishers[2]. The behaviour of the reader is therefore at least as important as the behaviour of the author in this regard. In the case of research publishing, the two populations may be substantially the same, but their behaviour can often be very different: for example, surveys have indicated that researchers, as readers, want papers to be as openly accessible as possible, but as authors are not willing to pay publication fees to make that happen (Nicholas *et al.*, 2005; Nicholas and Rowlands, 2005).

Similarly, while authors may be willing to archive their own work, it is possible that they may not be so willing to access other people's work outwith the channels of traditional journal publishing. However, Gadd *et al.* (2003b) indicates that, in fact, more researchers have accessed self-archived papers than have self-archived a paper themselves.

---

[1] RoMEO, a list of current publisher policies is maintained by SHERPA at http://www.sherpa.ac.uk/romeo.php?all=yes

[2] Early reports suggested that Google Scholar had patchy coverage and inaccurate citation records (e.g. Jacsó, 2005); however, more recent studies indicate that Google Scholar has overcome these early difficulties (Belew, 2005; Kousha and Thelwall, 2006). Google Scholar launched in 2004 and initially linked through to a full-text version, with alternative sources of the same paper also available as direct links on the main search page. However, with persuasion from journal publishers, this was changed so that the main link now always sends searchers to the official journal page even if they did not have access to the article (Pentz, 2005), while alternative versions have been moved to a separate page accessible from the main search results, meaning that searchers have to look harder to find different (self-archived) versions of papers.

It may take a while for readers to start using repositories to search for papers, as they may be used to the interfaces of their current search facilities provided by publishers. Kim (2005) observed users performing various tasks using two of the most-used institutional repository software programs, DSpace and ePrints, at Glasgow University, and found that neither were particularly easy to use. However, it is unlikely that researchers will often interface directly with a single institutional repository, as this would limit their search to papers by researchers at that institution. Rowland *et al.* (2004) recommends that content on individual repositories be centrally searchable, and many repositories follow the Open Archive Initiative standard of interoperability, meaning that they can be searched by external search facilities (Ware, 2004a). If the search facilities already used by researchers include self-archived papers, it is probable that they will be more likely to use these papers. Some general search facilities, such as Google, do include this material (Jacsó, 2003), and the question is whether these are facilities generally used by researchers in search of papers. Swan and Brown (2005) found that 71% of respondents to their survey used Google to find website-based articles, but this survey was conducted before the advent of Google Scholar, and researchers might now be using these tools more generally.

Even if researchers are aware of self-archived papers and where to find them, they might still choose to use the official published version because they feel it is more authentic and trustworthy. The main reason a reader might not feel able to trust a self-archived version is a fear that the paper has not been properly peer-reviewed (i.e. is a pre-publication version or has not been published in a journal at all) or that it differs significantly in other ways from the version available on the publisher's website or in the print journal.[3]

Wren's (2005) study of the impact factor of papers available on websites looked only at papers available as Portable Document Format (PDF) files. Wren's reasoning was that PDF papers are easier to archive by authors and preferred by readers, as the entire paper and its formatting information is contained within a single file. Certainly repositories tend to favour PDF as a format. However, there does not appear to be any firm data on the file format preference of researchers. It is likely that readers would view PDF files as more trustworthy, both for the reasons given above and because they are seen as less easily editable and more 'professional' in appearance. This should be especially true of the PDF file produced by the journal, or any PDF file which mimicked the journal's layout, as the reader would perceive the article to have the approval of the journal (and hence a reassurance of scientific quality) in a way that they might not with a file in the author's own layout (Antelman, 2006).

Motivations for authors

While there is a common perception amongst those authors who have never self-archived that the process is time-consuming and bothersome (Harnad and Brody, 2004), authors who have self-archived their papers by any means tend to find the process simple and quick. According to Swan and Brown

[3] Certainly it can be hard to tell whether a self-archived article has been peer-reviewed; in the very literature under consideration, Harris, 2005 criticises the CIBER report (Rowlands and Nicholas, 2005b) saying 'it is not clear that this study has been peer reviewed', while substantially the same report was published in Aslib Proceedings Rowlands and Nicholas, 2005a, a journal which reviews papers internally through its editorial board.

(2005), only 20% of authors self-archiving for the first time found the process difficult, and 54% said it was 'easy' or 'very easy'; two-thirds said that it took less than an hour. After the first time, the process was even easier and quicker: 54% of authors said that it took only a few minutes.

In addition to this subjective survey of authors' experiences, the amount of time it took to file an article in a repository was measured objectively by Carr and Harnad (2005). They found that for those authors who only deposited a few papers, the time taken to record all the necessary metadata in the archive was around ten minutes per paper, and this diminished as the author uploaded more papers (and presumably became more accustomed to the interface and the tasks required).

Depending on the terms and conditions imposed by the journal publisher, the self-archived article may or may not be identical to the version published in the journal, although it should be substantially the same. Some publishers refuse permission, at least for a period of time after initial publication in the journal, for the final peer-reviewed article to be published elsewhere (Gadd *et al*., 2003a). In this case, the author may only publish the article as submitted to the journal, i.e. before peer review has taken place. While in some very fast-moving fields these articles are valued, in many the lack of peer review is a major disadvantage.

Other publishers grant permission for the final version to be published elsewhere, but refuse to allow the archiving of the published PDF. In particular, the author generally does not have permission to use the journal layout and format, to which the publisher retains copyright regardless of copyright ownership of the article itself. This requires authors instead to archive their own copy of the final peer-reviewed article. In practice this means that the author may have to apply some corrections twice, particularly corrections made by copy-editors or by the author at proof stage. While these should not be substantial, the peer review having been already completed, the effect is to increase the effort involved for the author in self-archiving, or else have two different versions of the text available. It is possible that any difference between the published version and the self-archived version could undermine a reader's trust in the self-archived version, particularly as the publisher often requires the author to accompany the self-archived version with a statement such as 'The definitive version is available at [the publisher's website]', implying that the self-archived article is untrustworthy.

Nicholas and Rowlands, 2005 found that while many authors are indifferent to the possibility of there being several versions of their paper extant (and there could possibly be several; e.g. a preprint version, the journal published version, the author's own final version, and possibly a version with subsequent updates), a significant proportion are 'very unhappy' at the prospect. Authors also, according to this survey, tended to think it likely that institutional repositories would end up undermining the current system. Many authors who self-archive, however, use the journal PDF, regardless of whether they are permitted to do so (Antelman, 2006).

Current status

In 2004, the UK House of Commons Science and Technology Committee published a report on scientific publishing (House of Commons Science and Technology Committee, 2004). This report recommended that UK Research Councils and other Government bodies funding scientific research

mandate that results of work carried out under their funding be published in an open access form and that all research institutions establish open access repositories. The Government, in response to this report, decided that backing open access publications with funding or legislation would be unwise, preferring instead to let the market work unhindered (Shah, 2004). Harris (2005) comments that 'far from "facilitating" a level playing field, the Government seems to be hoping that, through inertia, the field will level itself'.

However, several UK funding agencies are now making it a condition of their funding that the results be published through an open access method. In particular, the UK Research Councils (RCUK) which between them account for a large proportion of academic research funding in the UK, have developed a policy of mandating self-archiving (RCUK, 2006). Meanwhile, the Wellcome Trust mandates open access publication and provides extra funds to cover the cost of publishing in author-pays journals (Wellcome Trust, 2005).

In the US, the National Institutes of Health has followed a similar line, with NIH-funded research to be deposited in the PubMed Central archive within one year of publication (National Institutes of Health, 2005). This constitutes a major proportion of US research in medicine and the life sciences. In order to avoid losing authors funded by these major bodies, publishers are being forced to change their conditions of publication to comply with these conditions (Wray, 2005). Moreover, the Wellcome Trust is working on a European equivalent of the US-based PubMed Central (Wellcome Trust, 2005), wherein all Wellcome Trust-funded authors would be required to place their articles within six months of publication. Similarly, RCUK are collaborating with the Joint Information Standards Council (JISC) to 'promote awareness of the existence and role of institutional and subject-based e-print repositories' (Wellcome Trust, 2005).

The other main source of research funding is the industrial sector. It is unlikely that commercial funders would be willing to see their funded work published on an open access basis. In some cases it might be considered to be good advertising for the company concerned, but for the most part we are unlikely to see a mandate for open access for research produced from this direction. The industrial sector as a source of readers is another matter, of course. This sector accounts for around 25% of papers written but nearer 75% of reading (King, 2004) and therefore it would be very much to the advantage of businesses to see academic papers published open access. Commercial funders are therefore more likely to encourage self-archiving as an option, as long as they were not expected to contribute towards the upkeep of repositories. However, as yet, the industrial sector does not appear to have made much contribution to the open access debate.

A large proportion of research articles have no specific funding (Rowlands and Nicholas, 2005b), particularly in medical and health-related fields. Again, for these papers there is no restriction on where they may be published, and no assistance for publishing in an author-pays journal. For these papers, then, it will be the authors' own choice where to publish, but as there is no extra funding, authors are more likely to choose a traditional journal, perhaps self-archiving the paper as well.

As more funding bodies encourage or mandate open access publication, journal publishers necessarily have to offer access policies compatible with those of the funders, or risk losing authors (and with them

not only copy but also potentially prestige and thus subscriptions) to other journals with more appropriate policies. Moreover, outwith Western Europe and North America open access publication is already held in much higher regard, with a much higher proportion of authors who have used some form of open access publishing and who think that open access will result in higher quality journals (Nicholas *et al.*, 2005). Journal publishers who wish to compete in the global market (particularly with regard to the fast-growing Asian research world) will have to work with these different attitudes.

Implications of open access

Goodman (2005) projected the potential impact of UK and US legislation, proposed at the time of writing, to mandate open access publication of results of all publicly funded scientific research performed in those countries. Goodman considered various scenarios relating to the form of open access publication that might be more generally adopted and whether traditional subscription-based publishing would be able to continue; he also included the case where the legislation was not enacted and open access publication remained voluntary. He considered that in all cases, even the non-mandatory case, open access publication would increase to over 95% of all primary research articles by 2020; he suggests that even without legislative pressures, other motivations for authors to publish their papers in an open access form will still drive the process forward, albeit more slowly.

Regazzi and Caliguiri (2006) indicate that institutional repositories are growing rapidly, while open access journals are not. Depending on whether or not users accept the self-archived versions as trustworthy, subscription-based journals may or may not survive a transition to mandatory open access publication. If libraries begin to cancel subscriptions because the articles are available freely online, publishers may be obliged to raise prices further, thus encouraging further cuts from the libraries, until the publishers are forced to discontinue publication of some journals.[4] Richardson (2005) suggests that availability of articles does reduce subscriptions, but Swan and Brown (2005) report that subscriptions to journals published by the Institute of Physics have not been affected by the presence of ArXiv. Goodman projects under this scenario that only 15% of existing subscription titles might still be published in 2013. If neither sufficient open access journals nor organised, peer-reviewed article databases are available, authors will publish instead by informal self-archiving without peer review.

This is, of course, a worst-case scenario. Goodman also examines the alternatives should subscription-based journals fail: that they be converted to or replaced by pay-to-publish open access journals (thus providing publishers with a continuing income stream), or that the repositories take over the functions of peer review. His favoured case is one where the publishers themselves take the lead in co-operating

[4] The Royal Society, in particular, issued a statement strongly opposed to current forms of open access publishing, both green and gold, for this reason (Royal Society, 2005), though they called for further research into the economics involved. However, they received adverse publicity for this stance (Macleod, 2005) when over forty Fellows of the Society wrote an open letter to the President of the Society requesting that he retract this statement (Ashburner *et al.*, 2005). The Society have since started a trial of an author-pays open access publication option in their journals (Royal Society, 2006), possibly because of fears that their journals would lose articles funded by the Wellcome Trust and the Research Councils if they did not. This example suggests that the growth of open access publication, in one form or another, and of a culture that encourages it is not something that publishers can ultimately avoid.

with institutions and with each other to provide open access repositories for articles published in their journals. This encourages authors to continue to submit articles to these journals, knowing that they will receive as wide an audience as possible, while the publishers maintain their income by 'making full use of opportunities for greater efficiency to lower prices' (thus encouraging libraries to subscribe to more journals) and discontinuing smaller journals altogether in favour of publishing their articles (fully peer-reviewed) only in repositories.

Goodman gives the example of the American Physical Society's price reduction as proof that such reductions are possible. However, Banks (2004) and Morris (2005) both point out that publishers have many overheads (e.g. supporting the journal editor, maintaining internet archives, managing the peer review process) and suggests that the cost to the publisher per paper produced is already greater than the cost to the library of purchasing it, the balance being made up by other income streams such as advertising and commercial reprints. These additional income streams would also be diminished by open access publication.

Baldwin (2004) gives a median figure of 15% (mean 18%) as the profit margin for a profitable society journal (one published by a learned society rather than a commercial publisher), bearing in mind that only two-thirds of the journals in the survey made a profit at all. Morris (2005) expands this figure to cover commercially-published journals as well.[5] Overall, then, there is not much room for publishers, particularly small society publishers, to make the sort of savings that Goodman suggests are possible.

Pinfield (2004) recommends instead that journal publishers become peer-review management operators, charging the institutions or authors in a similar way to author-pays open access publication. The Wellcome Trust believes that publishers will be able to survive a shift to an author-pays funding model without much difficulty (SQW, 2004). Costs of publication under this model will be lower, according to this analysis, as there will be no need for subscription management or licence negotiation, and fewer sales costs. Oxford University Press has found in its trial of an author-pays model of publication for *Nucleic Acids Research* that usage increased by 143%, and that though income per article declined, this was a continuation of a previously extant trend (Nicholas *et al*., 2006). For small societies with only one or two journals, however, these publication costs will not be substantially decreased, and many such societies, which use the income from subscriptions to their journals to subsidise their other activities, might be unable to cope with an author-pays model (Gannon, 2005).

[5] This does appear to conflict somewhat with Elsevier's 34% profit margin as quoted in the Science and Technology Select Committee's report House of Commons Science and Technology Committee, 2004. Reed Elsevier does not separate journals and books in its financial reports, so it is hard to tell for certain how this overall profit margin relates to Elsevier's journal sales, but in Reed Elsevier's 2004 Annual Report, journals in the Science and Technical division were said to be financially stronger than books. It is reasonable to assume that Elsevier, certainly the largest and most successful STM journal publisher with around a 45% market share in 2003, would have a higher profit margin than a small society publisher, and this means that the overall average profit margin for a journal should increase when commercial publishers are taken into account.

Questions remaining

Several studies over the past few years have shown, variously, that knowledge of and interest in open access is increasing amongst authors (Rowlands and Nicholas, 2005b); that authors are not willing to pay personally for open access publication (Rowlands *et al.*, 2004; Rowlands and Nicholas, 2005b) but think it more important that the journal is valued in the field (Schroter *et al.*, 2005); that awareness of self-archiving lags behind that of the author-pays model (Swan and Brown, 2005); and that though few authors currently self-archive (Swan and Brown, 2004a) most would be willing to in principle (Swan and Brown, 2005).

However, most of these surveys have concentrated in the main on the author-pays open access model. Much less research has been conducted into self-archiving and, in particular, little has directly studied the behaviour of readers of the self-archived material. Several studies have shown that citations of open access journals and self-archived material are increased (e.g. Harnad and Brody, 2004; Hajjem *et al.*, 2005), and this indirectly indicates that more people are reading the articles, but we do not have direct confirmation of this; in any case, most of these studies refer to open access journals rather than self-archiving. One recent study (Eysenbach, 2006) suggests that the citation advantage is less great for self-archived papers than for author-pays articles in the same journal, although this is based on a small sample size and needs further investigation.

There have been studies on user behaviour but these have mainly focussed on the behaviour of readers of electronic journals as opposed to print journals (e.g. Tenopir and King, 2002). Nevertheless, Tenopir *et al.* (2003) have shown that the proportion of papers accessed through non-journal means has increased to 35.8% amongst researchers in the 'advanced phase' of electronic publishing[6], and that within this there has been a large increase in the proportion of papers accessed as pre-prints. As repositories have developed, there has been a concomitant development of their use, from none in the 'intermediate phase' (normal online journals) to 10.2% of papers accessed in the 'advanced phase', while access of papers by inter-library loan or requesting a copy from the author has decreased. This implies that readers will use self-archived papers where they are given information about them, but does not answer the question of whether they will still make an effort to access the subscriber version and only use self-archived papers where they have no other choice. Moreover, it is worth investigating how 'advanced' other fields than physics and astrophysics have become since 2003; are researchers using independent journal search engine tools such as PubMed or Google Scholar which would give them information about articles hosted on author websites or in article repositories, or do they tend to restrict themselves to publishers' own search tools such as ScienceDirect or Blackwell Synergy?

In addition, most of these studies have been conducted using a population of corresponding authors of published papers. These are generally senior authors who are at the peak of their careers. While these are often the people who deal with publishing the paper they have co-authored, and who thus can be asked questions about their current publishing behaviour, for prediction of future behaviour it would also

[6] Tenopir *et al*. use the American Astronomical Society's advanced journal system – basically a fully featured article search engine – as an example of the advanced phase.

be informative to investigate the attitudes of more junior researchers. As they in turn become senior authors responsible for choosing the method of publication of their papers, their views now will have greater impact in years to come.

Methods

Rationale

I intended to explore the reading behaviour and attitudes of academic researchers and end users, and to correlate it to the archiving behaviour and attitudes of the same population as authors. To do this, I surveyed a sample of UK academic researchers using a web-based questionnaire. I chose a web-based format for the following reasons:

- all universities are highly internet-connected, and so it can be assumed that a good sample of academic researchers at all levels can be contacted this way
- it is cheaper to send a link to the questionnaire through email than to send a paper questionnaire through the post, and so a larger initial sample size can be obtained
- email addresses for staff at UK academic institutions can be easily obtained through their institutions' websites
- as it is easy for the recipients of the email to click through and start the questionnaire, there is likely to be a higher response rate than there would be for a paper questionnaire
- the questionnaire can be tailored so that questions are presented based on the answers to previous questions, thus ensuring that respondents are not presented with irrelevant questions.

The questionnaire can be seen in Appendix 1.

Aims and hypotheses

The survey focuses specifically on the behaviour of the population with respect to self-archiving, primarily through institutional and subject-based repositories but also in terms of personal websites. The aims of the questionnaire were to determine:

1. whether awareness of repositories and repository attitudes and behaviour differ between established and new researchers, and between research fields
2. whether the knowledge and behaviour of readers who are also authors differs from that of readers who are not also authors
3. whether readers view the journal PDF as more authoritative than other formats, and whether they view journal websites as more authoritative article sources than repositories
4. whether readers primarily source papers using facilities that enable them to search repositories, or whether their sources do not include them

5. whether self-archived versions are indeed as authoritative as the journal PDF (i.e. whether they have essentially the same content as the journal version and are recognisable as being a peer-reviewed article from a specific journal).

My hypotheses were as follows:

1. Junior researchers are more willing to use self-archived papers than senior researchers, and use of self-archived papers is only common in those fields where self-archiving is widespread.

2. End-users (non-authors) are unaware of archives; researchers are unaware of institutional archives even within their own institution; awareness of subject repositories is more widespread, but depends on the subject.

3. Users prefer the journal PDF, sourced from the journal website or publisher search engine, and only search for other versions if the journal PDF is not available to them; many users will simply give up if they cannot access the journal version.

4. Nevertheless, many researchers use search tools which include repository articles, and may be accessing self-archived material without being aware of it.

5. Most papers are not self-archived; most authors who do self-archive upload the same version that appears in the journal regardless of whether they are permitted to do so, thus preserving the authentic and authoritative appearance of the journal article.

Questionnaire design

Demographics

Questions 1–6 and 18 were designed to gather basic data on respondents, including their institution, subject area, seniority as measured by role and length of time in the field, and activity as measured by article use and article publication (Aim 1). I did not feel it necessary to ask for personal data such as age and sex, and felt that this might discourage people from completing the questionnaire.

Search engine use

Questions 7 and 8 were designed to determine whether researchers would naturally come across self-archived articles during the course of their search, or whether their search was limited to journal-hosted versions of papers (Aim 4). Of the search facilities listed in the questionnaire, some, such as Emerald, Synergy, and Metapress, are run by publishers and thus provide only official journal versions to subscribers; others, such as OAIster, are designed specifically to search repositories. Facilities such as Google and Google Scholar search both journal sites and repositories, as well as authors' personal websites. The question also included several of the major subject-based repositories as search options, including ArXiv (physics and astrophysics), Cogprints (cognitive science), and PubMed, a search facility which links to the National Institutes of Health repository, PubMed Central (medicine). I asked respondents to choose all the facilities that they used, and also asked them to choose their preferred

facility. I also provided respondents with an option to choose 'other' and add any additional facilities that they used.

Questions 9 and 10 dealt with acceptable and preferred paper formats of articles downloaded by respondents, with the expectation that respondents would prefer non-editable formats of the sort typically made available by publishers as 'official' versions (Aim 3).

Repository attitudes and use

Questions 11–13 covered respondents' knowledge and valuing of self-archiving. I gave a brief definition of self-archiving and asked them to state their level of knowledge. I did not wish to give too much information for fear of biasing the results. Question 12 asked respondents to agree or disagree with statements designed to reveal how accurate their beliefs were (Aims 1, 2 and 3). Question 13 asked respondents to state whether their institution had a repository, and this was expected to tie in to knowledge of self-archiving (in order to check the accuracy of respondents' answers to this question, as well as to investigate the effect of the presence of an institutional repository in general, I investigated each institution represented in the survey; see Appendix 2). Questions 14–16 dealt with respondents' use of repositories both institutional and subject-based as well as personal websites, asking respondents to state whether they had used them under a number of different conditions (Aims 1, 2, 3 and 4). Question 17 explored respondents' attitudes to self-archiving by offering ten statements with a five-point scale of disagreement–agreement for each one (Aim 3).

Author behaviour

As not all respondents were expected to be authors of papers, the results of Question 18 ('In what year was your last paper published?') were used to determine whether the remaining questions concerning author behaviour were presented. If the respondent replied 'I have never been an author of a published paper', the author behaviour questions were skipped and the respondent was presented only with the space for comments and opportunity to provide an email address for future correspondence.

Question 19 was intended to give another measure of the seniority and research activity of respondents by the number of papers they had published in the past year (senior researchers would be expected to publish more papers). Question 20 asked for the name of the journal that the most recently published paper was submitted to, in order to check the accuracy of respondents' answers to the following questions (21–23) about the conditions set by the publisher on self-archiving of their paper (Aim 5). Questions 24–28 dealt with the respondents' intention to self-archive and the reasons they came to their decision (Aim 1) and the version and format they chose to self-archive (Aim 5). In addition to the multiple-choice options covering what I expected to be the most common answers, I also gave the option for respondents to add comments and additional answers to all of these last questions.

Pilot

I tested the questionnaire with a small pilot group of six people, which included academics and professionals in the fields of psychology, medicine, arts and media. I incorporated their suggestions and feedback into the questionnaire.

Sample selection

Email addresses of UK academic researchers were collected from university websites by David Williams, Ben Williams and Robin Miller. In total, the invitation email was sent to 2775 academic researchers at all levels (from PhD student to professor) and across all fields. Email addresses were collected from a random selection of departments in each of the following universities:

- University of Aberdeen
- Cardiff University
- Keele University
- University of Lincoln
- Reading University
- University of Southampton
- University of Surrey
- University College, London
- University of East Anglia

These universities were chosen to provide a balance between insititutions with and without institutional repositories. I was particularly keen to include the University of Southampton because of its history of activity in the field of self-archiving, with a highly populated institutional repository and (in one department) a mandatory self-archiving policy: this represented a population who would be expected to know more than the average about self-archiving. The other universities were chosen solely on the basis of the presence or absence of an institutional repository, without further knowledge about the repository's status.

In addition, email addresses of a small population of 100 UK health professionals were provided by Dr Elizabeth Evans (Tudor Gate Surgery, Abergavenny). These represented an additional population of non-author end-users.

The questionnaire was available at http://www.tardis.ed.ac.uk/~rhi/Survey1.php from 15 July, and invitations were sent by email from 04008503@napier.ac.uk on 27 July. Responses were automatically loaded into a database through FormTools (http://www.formtools.org), and I downloaded and analysed all the responses that had been gathered by 29 August. I analysed the data using SPSS version 14 (SPSS Inc., Chicago, IL) and Microsoft Excel (Microsoft, Redmond, WA).

Results and discussion

Overview

Over four weeks, I received 438 unique responses to the questionnaire, from 2,775 invitations to academics and 94 to medical professionals. Ninety-six messages were returned as undelivered. This therefore corresponds to a response rate of 16%. Several people also sent replies to the invitation email saying that they did not complete the survey because they were not researchers and/or did not use papers.

I also received two replies to the invitation email from people who had started the questionnaire, but felt that they could not complete it: because they knew little about self-archiving, they felt that this survey was not 'for them'. (Others who did complete, or partially complete, the survey added comments to the same effect.) This is unfortunate, as one of the purposes of the survey was to assess respondents' level of knowledge about self-archiving. It can be assumed that these emails are representative of a rather larger population of potential respondents who felt unable to complete the questionnaire for the same reason. These results therefore need to be read in the knowledge that the sample may be somewhat biased in favour of people who already know something about self-archiving, and are more likely to have used self-archived papers.

It must also be borne in mind that the sample size is too small to be truly representative. Some subject areas, for example, are represented by fewer than ten respondents, and so it is risky to generalise too much from these data. However, these results can still give an indication of trends and attitudes that can be further explored in future, more extensive studies.

Sample demographics

Institution and repository status

Institution	Frequency
Aberdeen	32
ACM	1
Bournemouth	1
Bristol	1
Cancer Research Institute	2
Cardiff	53
Durham	1
Edinburgh	1
Eurecom	1
Glasgow	1
John Innes	1
Keele	12
Lincoln	23
Macaulay Institute	1
Macerata	1
Middlesex	1
MRC	3
NASA	1
Newcastle	1
NHS	10
NICTA	1
Oxford	1
Reading	35
Robert Gordon	1
Roehampton	15
Rutherford Appleton	2
Southampton	81
Staffordshire	1
Surrey	5
Thessaloniki	1
UCL	42
UEA	66
Total	399
Did not answer or unidentifiable	44
Total	443

Table 1: Institutions of respondents

Table 1 shows the institutions respondents were affiliated to. This is a much greater range of institutions than that to which the emails were originally sent. It was to be expected that some respondents would

be affiliated to some other institution as their main institution and only secondarily to the institution at which they received the email, and this was acknowledged in the questionnaire. About 10% of respondents did not answer this question, or their answers were unclear and the institution could not be traced.

I checked each institution against the repository databases at ROAR (Registry of Open Access Repositories; http://archives.eprints.org) and OpenDOAR (Directory of Open Access Repositories; http://www.opendoar.org), and obtained additional information about each institution's repository, where present, from individual institutional repository websites. If neither the databases nor a search of the institution's own website turned up a respository for that institution, I assumed that it did not have one. Where contact details were given on an institution's repository website, I emailed the repository maintainers with additional questions (see Appendix 2 for more details on each institution's repository). This resulted in the analysis of institutional repository status for these institutions shown in Table 2.

Institution	Actual repository status
Aberdeen	yes
ACM	no
Bournemouth	no
Bristol	yes
Cancer Research Institute	yes
Cardiff	yes
Durham	yes
Edinburgh	yes
Eurecom	no
Glasgow	yes
John Innes	no
Keele	no
Lincoln	yes
Macaulay Institute	no
Macerata	no
Middlesex	yes
MRC	no
NASA	yes
Newcastle	yes
NHS	no
NICTA	no
Oxford	yes
Reading	no
Robert Gordon	no
Roehampton	no
Rutherford Appleton	yes
Southampton	yes
Staffordshire	no
Surrey	yes
Thessaloniki	yes
UCL	yes
UEA	no

Table 2 Institutional repository status

Just over half of the institutions in this study had an institutional repository. It should also be borne in mind that some institutions, such as UEA and Roehampton, shown in this table as not having repositories are in fact actively developing them (see Appendix 2). However, since they are not yet fully released, for the purposes of this study I decided not to consider these repositories. Moreover, since the institutions with repositories tend to be the larger institutions (SHERPA, 2006), it is unsurprising that a disproportionately high proportion of respondents (62.2% of those whose institution could be traced) belonged to such institutions.

I aggregated those institutions with fewer than five respondents into two categories, depending on whether or not the institution had an institutional repository. The resulting set of institutions is shown in Figure 1.

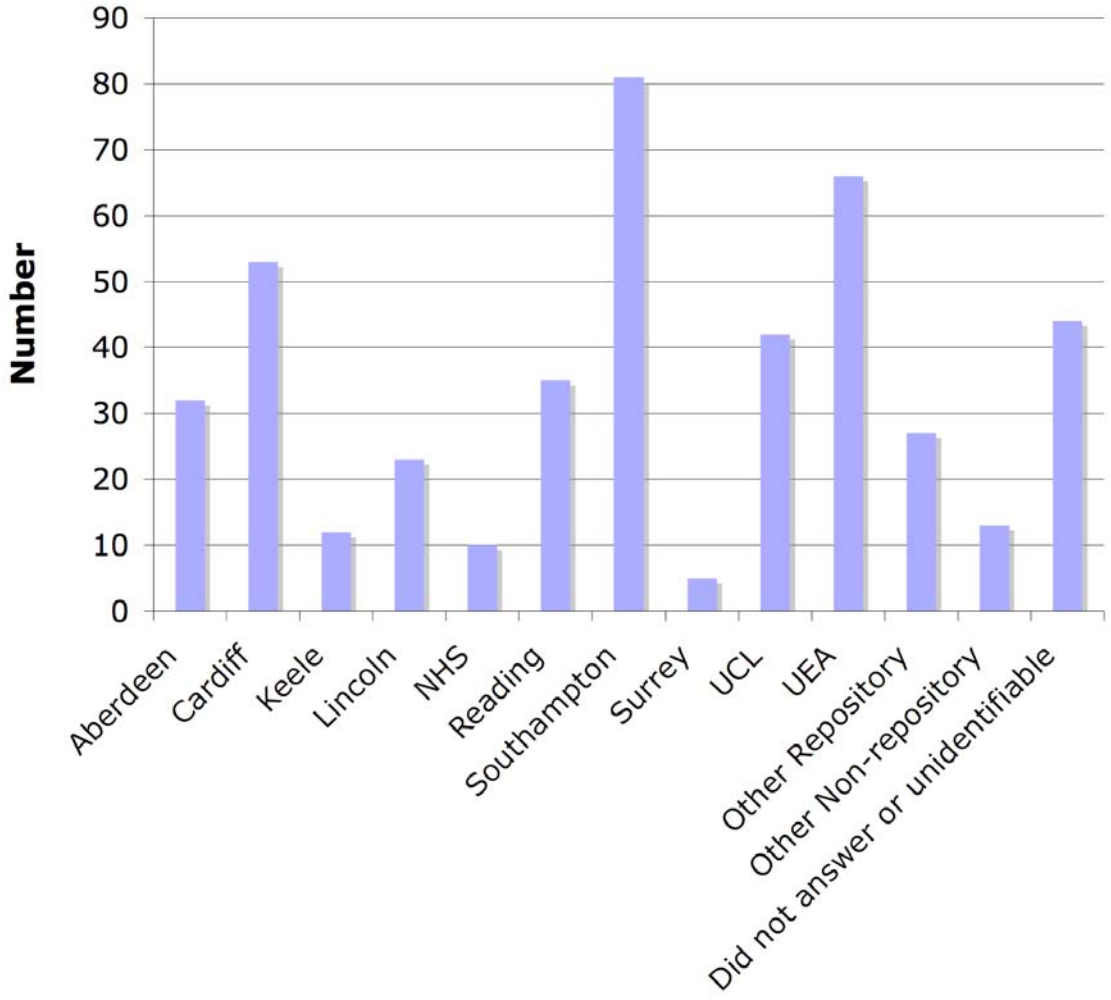


Figure 1: Numbers in the institutions (aggregated version)

The total number of respondents from institutions with and without repositories is thus as shown in Figure 2.

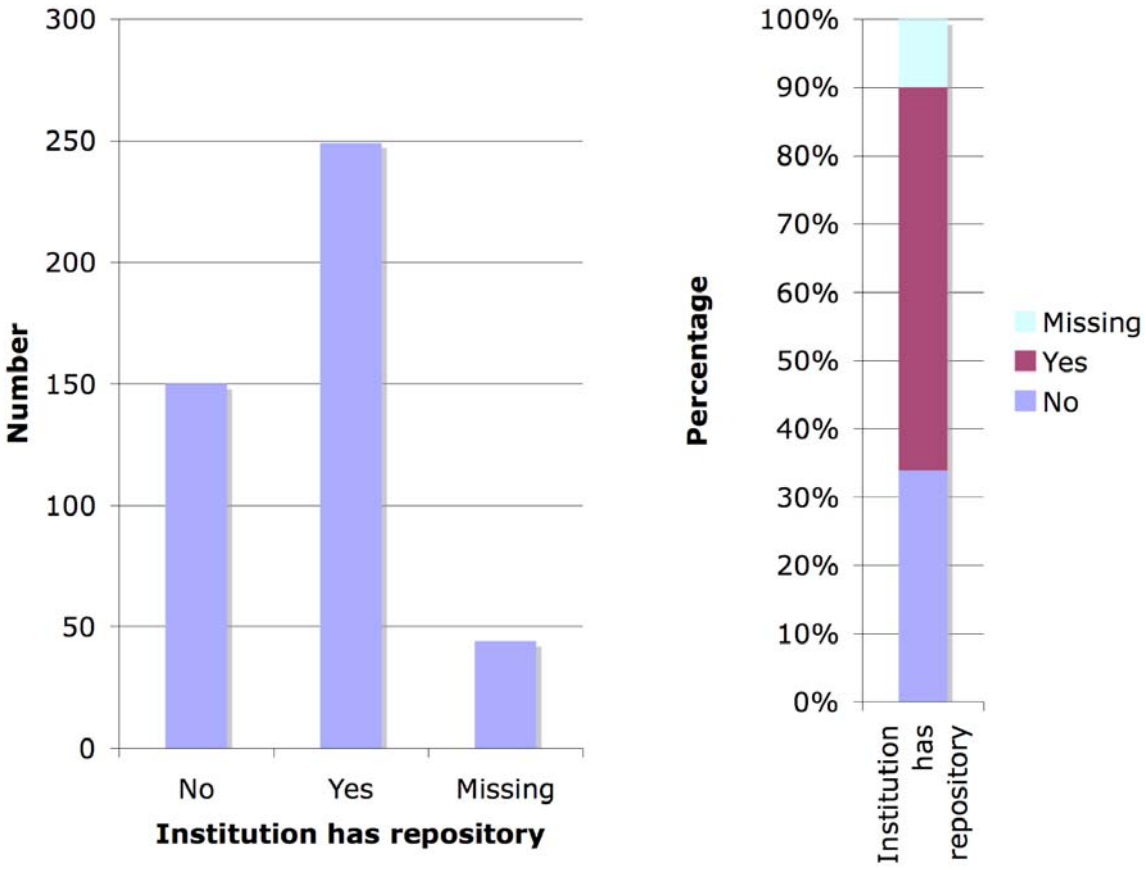


Figure 2: Institutional repository status

As can be seen from Figure 2, of those respondents whose institution can be identified, just under two thirds come from an institution with a repository. As mentioned above, this is likely to be because the institutions with repositories tend to be larger, and so this is likely to reflect the population of researchers at large. Also, it is possible that respondents working where there is a repository are more likely to have completed the survey because they were more aware of self-archiving; the predominance of respondents from the University of Southampton (Figure 1), which has the most active and highly publicised repository (see Appendix 2) indicates that this is the case.

Question 2 was originally intended to distinguish between academic and commercial researchers; however, in the end, the invitation was only sent to university departments (and a small number of health professionals) and not to commercial establishments. As all but 21 of the 433 respondents who answered this question chose 'Academic', I did not use this variable in the analysis.

Subject field

The distribution of subjects (see Figure 3) was heavily weighted in favour of the sciences. This partly reflects the actual population, as research departments in the sciences tend to be much larger than those in the arts or humanities. Because of the low response rates for these categories, I aggregated 'Arts', 'Humanities' and 'Languages' together into one 'Arts and Humanities' category. In addition, and

despite the high response rates of scientists in general, very few physicists responded to the questionnaire. The category of Physics/Astronomy was therefore aggregated with the Engineering category.

Those who responded 'other' and stated their field were classed in the most suitable existing category (e.g. 'Nursing' was categorised as 'Medicine'); however, a sufficient number of people added 'Environmental Science' that it was worth creating a new category for them (Figure 3).

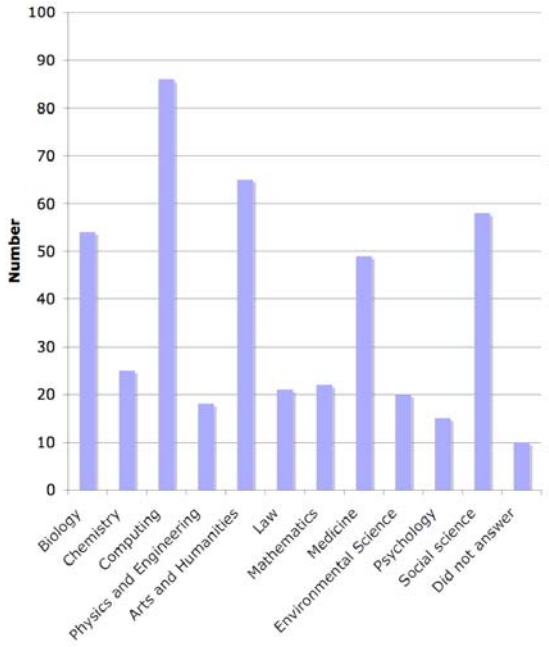


Figure 3: Numbers in the various subject fields

A consideration worth investigating is the possibility that analyses involving the subject area might be confounded by the institution. This is possible because of the way the sample was collected, mostly from university departmental websites, which might result in a cluster of respondents from the same department skewing the results for a given institution or subject. To investigate this, I analysed the relationship between subject area and the presence or absence of an institutional repository (Figure 4).

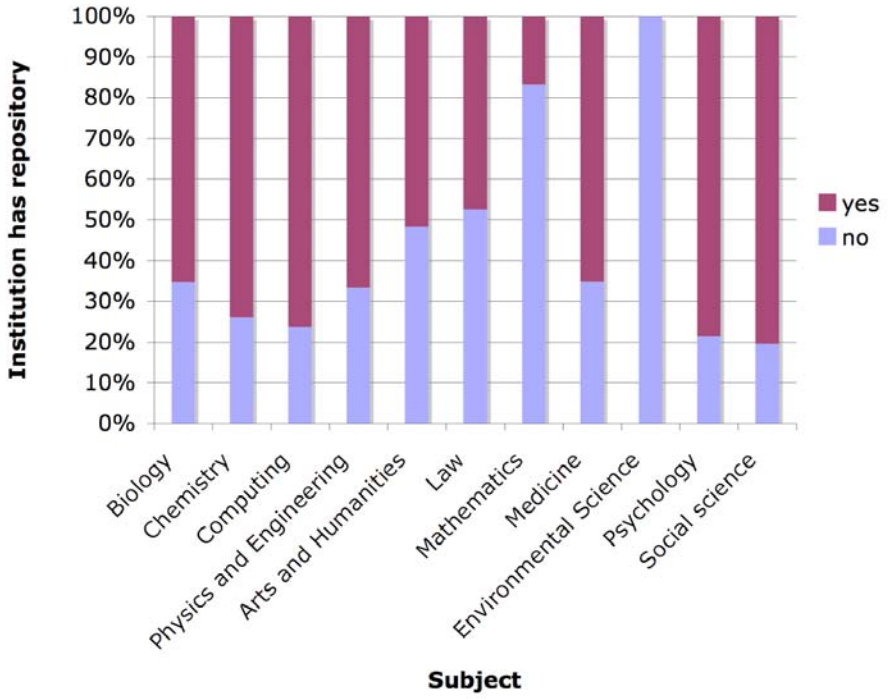


Figure 4: Institutional repository presence by subject field

There were significant differences (χ^2=67.8 df=10, p<0.001) between subject areas: scientists were generally more likely to work in institutions with repositories, and arts and humanities researchers tended to work in institutions without repositories. This could be a coincidence caused by the small sample size and the way the sample was chosen, or it could be that scientists tend to work in larger institutions which are more likely to have repositories. In particular, all the Environmental Science respondents came from institutions without repositories, and in fact almost all of them came from a single university. These respondents may therefore not be entirely representative of environmental scientists in general. It should be borne in mind, then, that any observed effects involving primarily Environmental Science in this study could be due to either the institution or the subject. It should also be noted that the relatively small number (22) of mathematicians mostly come from institutions without repositories, and the results should be read with this in mind.

Role and number of years in field

These were intended to give a measure of the seniority of respondents. The numbers of respondents in each role are shown in Figure 5.

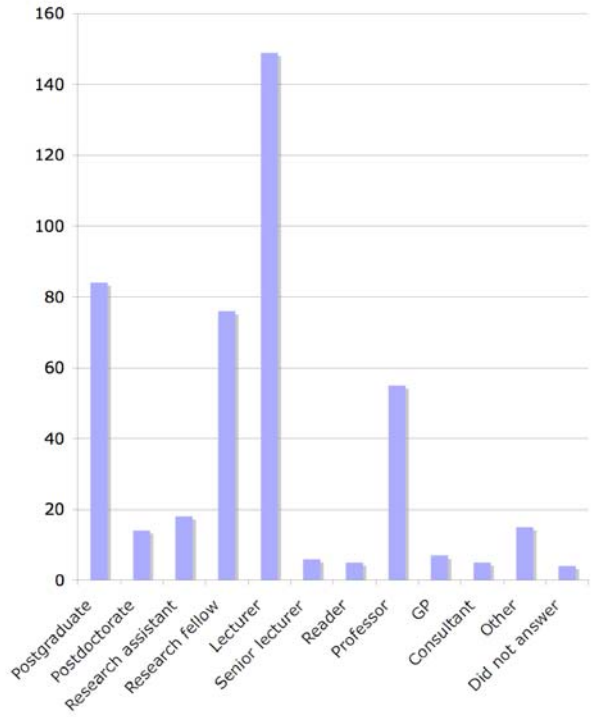


Figure 5: Numbers in the various roles

There is a considerable bias towards younger researchers. This was encouraging, as previous studies have tended to concentrate only on senior authors of papers. The attitudes and behaviour of younger researchers is important in helping to judge future trends, as these are the people who will become senior authors in a few years' time. However, it is possible that there may be some error here, as the positioning of this question directly after the question on role may have led some people to believe that what they were being asked was the length of time that they had been in their current *role*, rather than in their field as a whole.

The length of time respondents had spent in their respective fields is given in Figure 6. Assuming that most respondents have not recently transferred from another field, this should give some indication of the seniority of respondents. There is a good spread within the categories, and a bias towards those who have been in their field for a shorter time. The two measures correlate reasonably well (Figure 7; χ^2=285.7, df=40, p<0.001), indicating that, except in a few cases, respondents were interpreting the questionnaire correctly and listing the length of time they had been in the field as a whole rather than in their current role.

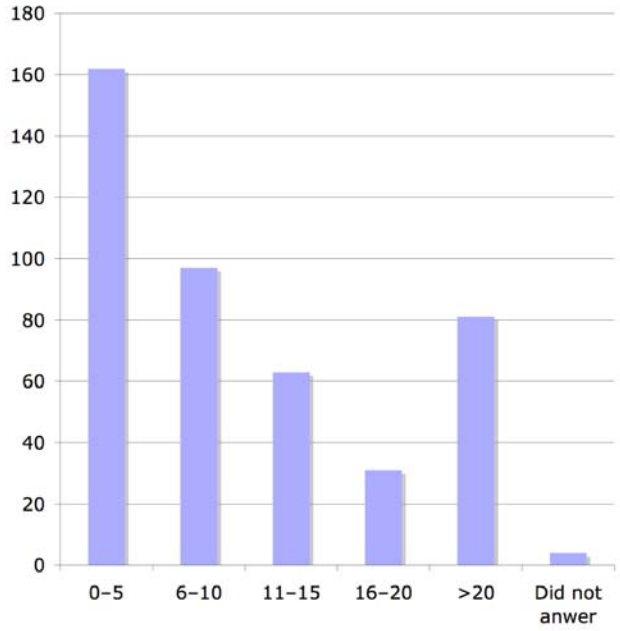


Figure 6: Number of years in field

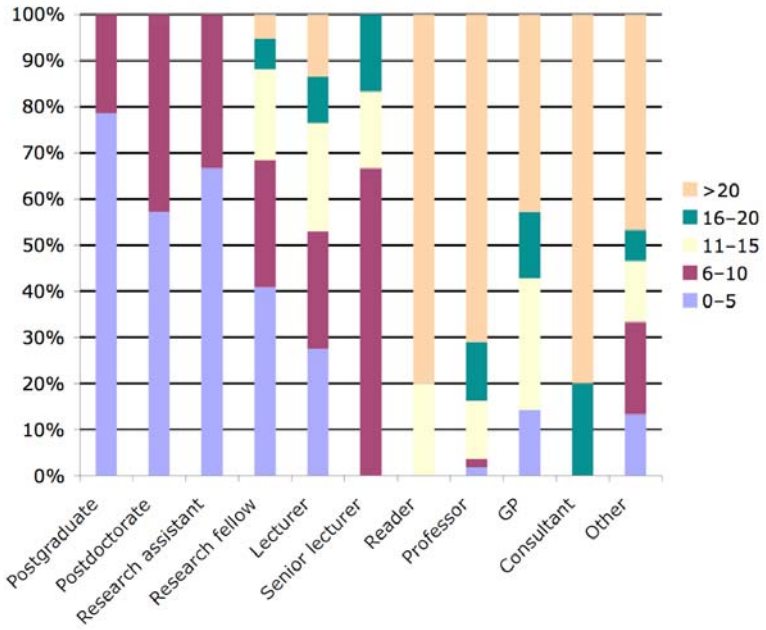


Figure 7: Relationship between role and time spent in field

The length of time that respondents had been working depended to some extent on the subject area they were working in, as can be seen in Figure 8.

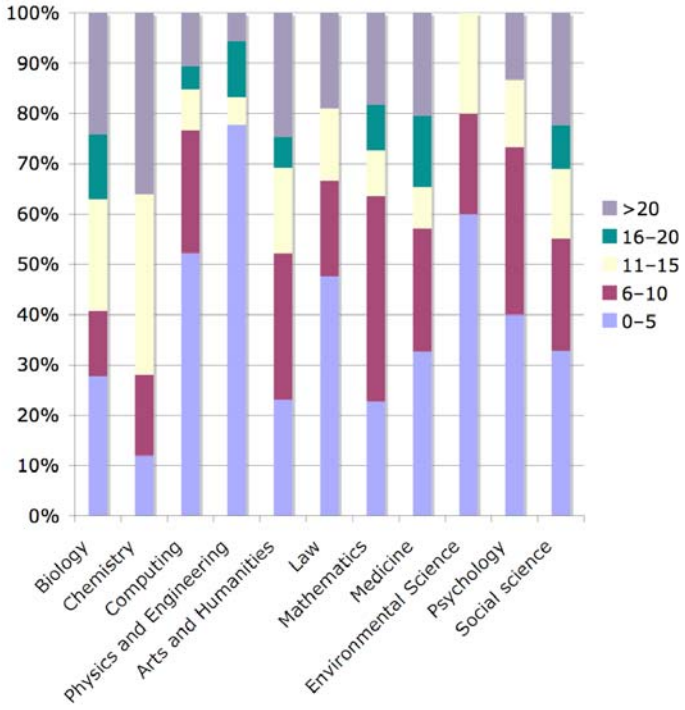


Figure 8: Subject area and years spent working in it

Again, this may partly be a feature of the way the sample was chosen and could be affected by the low response rate, but it is noticeable that the more 'traditional' subjects have a broader and more even spread.

Number of papers used per year

This gives an indication of the activity of respondents and how much use they make of literature in their field. It also suggests the frequency with which they search for articles.

Again, there is a good balance across the categories; the 0–10 category has fewer members only because it is a smaller category by definition. I felt it important to distinguish between those who use very few papers and those who are more active researchers.

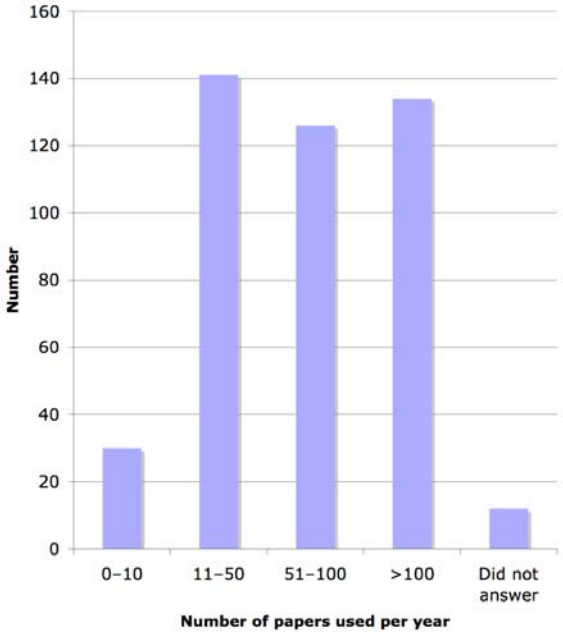


Figure 9: Number of papers used per year

The subject field in which respondents work would be expected to have an effect on the number of papers used, and this is indeed the case, with respondents in the field of mathematics in particular using fewer papers than those in other subjects (Figure 10; χ^2=85.0. df=44, p<0.001).

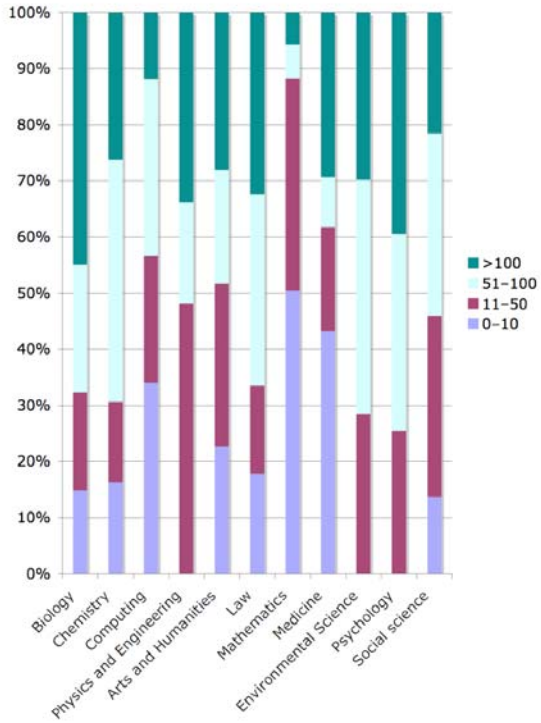


Figure 10: Relationship between subject area and number of papers used per year

Number of papers authored in the past year

Eighty-one per cent of respondents (350) had published a paper at some point. For these respondents, the number of papers each respondent had written in the past year is given in Figure 11. Thirty-nine had not published a paper in this time, while 13 had written more than ten. Most respondents had written between two and five papers. There was no significant relationship between this and subject area, institution, or seniority as measured by role or number of years in the field.

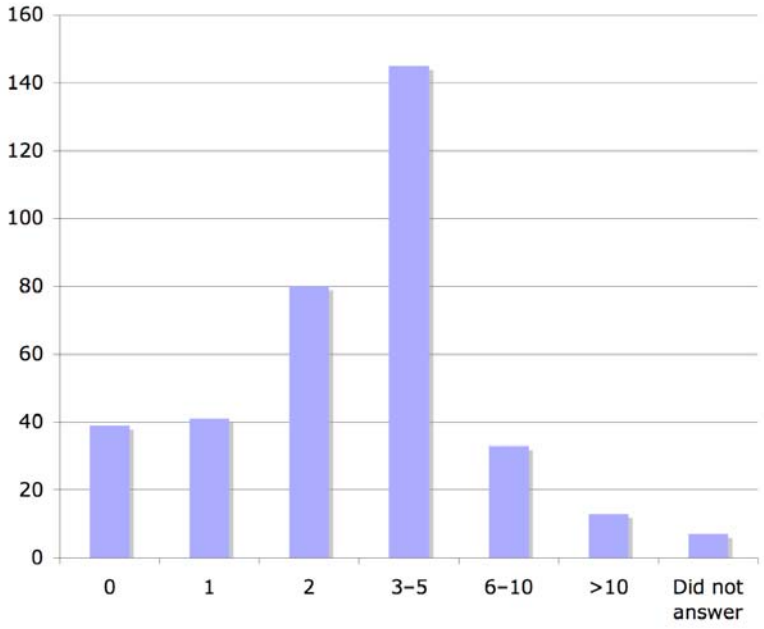


Figure 11: Number of papers published by authors in past year

Searching behaviour

Database/search engine use

Question 7, 'Where do you search for articles online' was designed to uncover whether respondents were already using sources of information that included repositories, or whether they mainly searched publisher-based databases that only presented the journal (subscriber-limited) version. Accordingly, the list presented in the questionnaire included publisher websites, repositories, and independent search engines and abstracts databases that included one or both sources of articles.

Table 3 shows a breakdown of each item in the list by the source of the data it presents to searchers. For the purposes of this analysis, I assumed that internal institutional search facilities gave access to both publisher-based and repository-based data, on the grounds that, for those institutions with repositories, the internal search facility would probably offer the service of searching the institution's own repository; even where the institution has no formal repository it might still be possible to search some of the institution's own research output. Of course, it is difficult to check this because most such facilities are restricted to certain networks or are password-protected.

	Publisher-owned	Publishers	Repositories	Web
ArXiv			Yes	
ACM	Yes	Yes		
ADS		Yes	Yes	
ASSIA		Yes		
BIDS	Yes	Yes		
Biomed Central	Yes			
		Yes		
CINAHL		Yes		
CiteBase			Yes	
CiteSeer		Yes	Yes	
Cogprints			Yes	
CSA		Yes		
EBSCO		Yes		
Edina		Yes		
Embase	Yes	Yes		
Emerald	Yes	Yes		
Engineering Village		Yes		
FAME	Yes	Yes		
FindArticles		Yes		
Google		Yes	Yes	Yes
Google Books		Yes		
Google Scholar		Yes	Yes	
HeinOnline		Yes		
IBSS		Yes		
IEEE	Yes	Yes		
Ingenta	Yes	Yes		
Internal institutional		Yes	Yes	
ISI		Yes		
JSTOR		Yes		
Lexis	Yes	Yes	Yes	
LISA	Yes	Yes		
MathSciNet		Yes		
Medline		Yes		
Metapress	Yes	Yes		
MSN				Yes
OAIster			Yes	
Ovid[7]		Yes		
Proquest		Yes		

[7] Ovid offers links to full-text articles in open access journals, and includes articles in PubMedCentral, which is a repository and includes author manuscripts for non-full-text journals. However as this is such a minor part of Ovid I have not considered it to count as searching repositories.

PsycINFO		Yes		
Pubmed		Yes	Yes	
RePEc		Yes	Yes	
SAGE	Yes	Yes		
ScienceDirect	Yes			
		Yes		
Scirus	Yes	Yes	Yes	Yes
Specific journal website	Yes			
		Yes		
SpringerLink	Yes	Yes		
Synergy	Yes	Yes		
Westlaw		Yes	Yes	
WilsonWeb	Yes	Yes		
Yahoo				Yes

Table 3: Search engine and database properties

Respondents were asked to select all the online search facilities that they used, and were also given the option of specifying further facilities in a free-text box. Many respondents made use of this option, and provided a wide range of additional search facilities. Some of these were different names for the ones already offered (offering slightly different services but searching the same databases, e.g. Pubcrawler for PubMed[8] and Scopus for Scirus), and I aggregated these with the appropriate entry in the questionnaire. Where five or more respondents entered the same facility, I created a new category for that facility; otherwise, the facility was classed as 'Other publisher', meaning that it was publisher-based or searched only publisher-based data; 'Other repository', meaning that this facility was a repository or searched only repository-based data; or 'Other both', meaning that it searched both publisher-based and repository-based data.

The frequency with which each search facility was chosen is given in Table 4.

	Frequency
Google	283
Internal institutional	253
Specific journal website	238
Google Scholar	211
Medline/PubMed	195
ScienceDirect	178
Ingenta	140
SpringerLink	91
CiteSeer	65
ISI	63
Other Publisher	55
Ovid	51
EBSCO	48
Synergy	45

[8] I also aggregated PubMed and Medline, as the Medline database constitutes the great majority of articles listed in PubMed.

SAGE	36
Engineering Village	30
Scirus	29
Biomed Central	27
Proquest	23
Emerald	22
Metapress	22
IEEE	22
Other repository	20
Embase	19
Yahoo	19
Edina	14
FindArticles	14
Google Books	13
Westlaw	12
JSTOR	12
ACM	10
MathSciNet	9
Other Both	9
CiteBase	8
ASSIA	7
BIDS	7
PsycINFO	7
Lexis	7
IBSS	5
CINAHL	5
HeinOnline	5
ArXiv	5
FAME	4
RePEc	3
Cogprints	3
LISA	3
WilsonWeb	3
MSN	3
ADS	2
CSA	2
OAIster	0

Table 4: Choice of search engines

The strong showing of Google is particularly notable. As a web search engine, it would pick up all types of article, whether on publishers' websites, web-based repositories, or personal web pages. Google is particularly well known and easy to find, and its popularity may have something to do with this. A smaller but still high number used the more specialist Google Scholar. Swan and Brown (2005) also found that a high proportion of researchers used Google to search for papers, although this study was conducted before the launch of Google Scholar.

I then analysed the selection of facilities chosen by each respondent to determine whether, overall, they used facilities which gave them access to only publisher-based articles, only repository-based articles, or both. Because of the potential variation in internal institutional search facilities, I excluded these from the analysis, choosing 'Repository', 'Publisher', or 'Both' based only on the respondents' other choices. The results are shown in Figure 12.

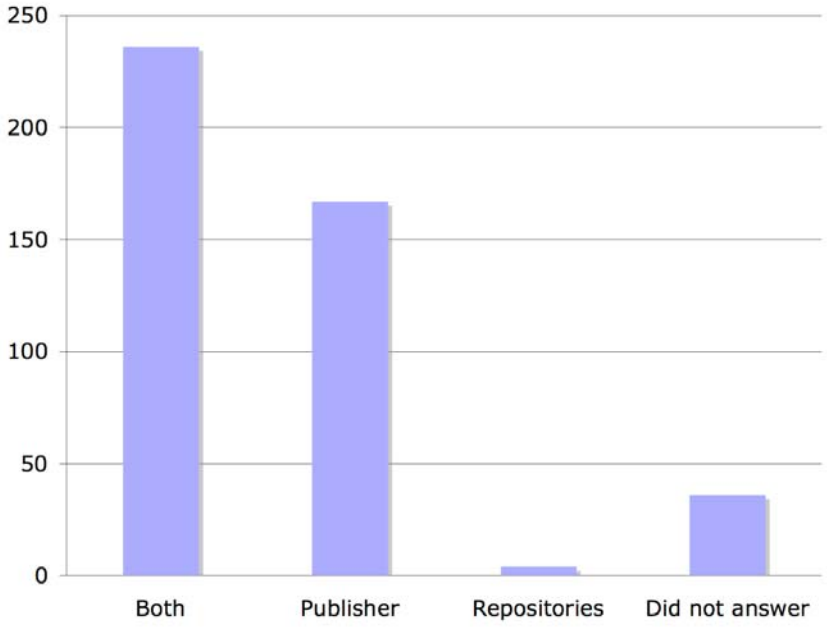


Figure 12: Overall properties of search facilities used

A majority of respondents made use of search facilities that offered them access to both publisher-based and repository-based articles.

I also asked respondents to choose their preferred facility with the question 'Which **one** of these do you use the most?' Respondents were offered the same list as in the previous question, including an 'Other' option. Where a respondent chose 'Other' and had listed several facilities in the free-text box, I took the first one to be the preferred option. The preferred search facilities are shown in Table 5.

	Frequency
Google	84
Internal	67
PubMed/Medline	54
Google Scholar	38
ISI	35
ScienceDirect	26
Journal website	17
Ovid	12
Other Publisher	11
MathSciNet	8
Ingenta	7
IEEE	6
ACM	5
CiteSeer	5
Scirus	4
BIDS	3
Emerald	3
Other Both	3

Edina	2
Embase	2
Google Books	2
ADS	1
CINAHL	1
CiteBase	1
CSA	1
EBSCO	1
Engineering Village	1
IBSS	1
OAIster	1
Proquest	1
SAGE	1
Synergy	1
Other Repositories	1
ArXiv	1
Total	407
Did not answer	36
Total	443

Table 5: Preferred search facility

Again, Google predominates, showing that many are using this as their main means of finding papers. This was not limited to young researchers; there was no statistically significant relationship between the number of years spent working in the subject field and the preferred search facility. However, it did depend on the field of research: computing researchers were over twice as likely as expected to use Google as their preferred means of finding papers, and these accounted for 38 of the 84 respondents who chose Google as their most favoured means of finding papers online.

An analysis of preferred search facilities based on their categorisation in the same manner as in Figure 12 gives results as shown in Figure 13.

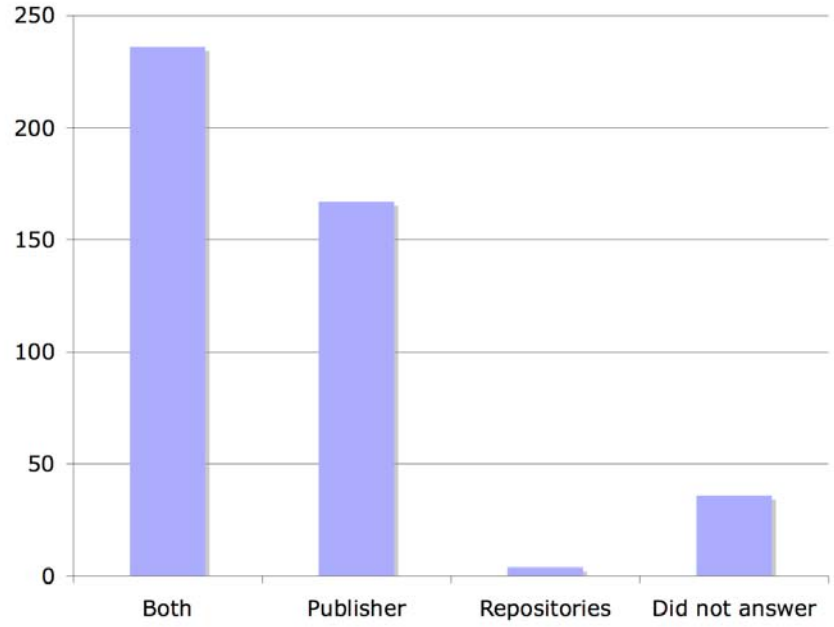


Figure 13: Preferred search facility properties

Over 50% of respondents are using, as their preferred search facility, a service that gives access to both publisher-based and repository articles. This appears to show a high preference for such facilities. However, this is affected by the predominance of internal institutional search facilities. As mentioned above, it is difficult to tell whether these actually search repositories or not, as this will vary between institutions. Figure 14 shows the results with the internal institutional search facilities shown separately.

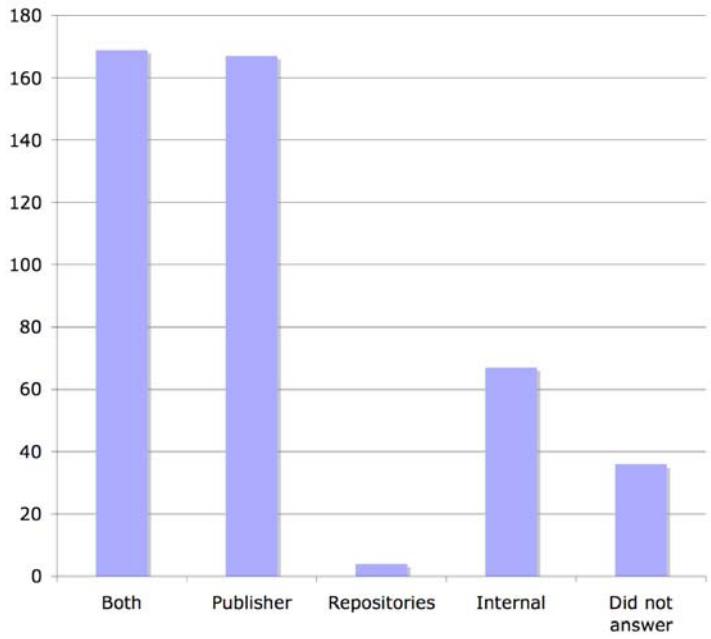


Figure 14: Preferred search facility properties ('Internal' not defined)

This shows that about equal numbers of respondents prefer publisher-based search facilities and facilities that give results from both publisher databases and repositories. Hardly any respondents preferred facilities that were entirely based on repositories; and in fact one of these may be an error, as one respondent chose OAIster as their preferred facility but did not select it in the preceding list of all search facilities used.

Preferred paper formats

Questions 9 and 10 dealt with the formats in which people prefer to read articles they access online. A large number of respondents chose an option for 'most preferred format' in Question 10 which was not any of the formats they had chosen as 'preferred formats' in Question 9. This was particularly noticeable with PDF, which hardly any respondents selected as one of their preferred formats, yet a huge majority selected as the format they most preferred. I assumed that the most preferred format should also be included in the list of preferred formats, and that the question design was probably to blame (in Question 10 'Of these, which one do you prefer the most', the term 'Of these' was intended to refer to the list in the preceding question, but could have been taken to mean the list given in Question 10; the two lists could therefore have been interpreted as being mutually exclusive).

Assuming the above, then, the frequency of choice of each format as an acceptable format is shown in Figure 15.

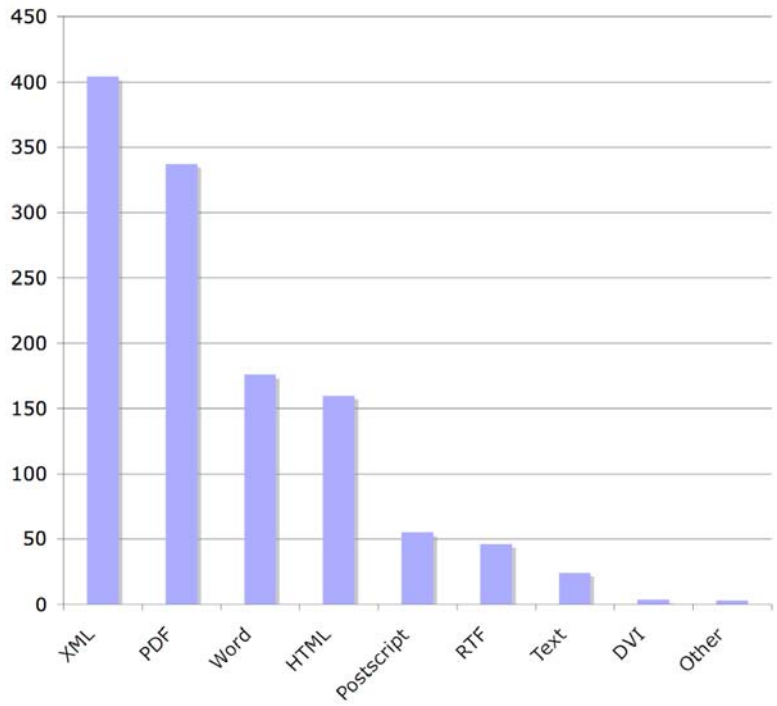


Figure 15: Choice of online paper formats

XML is the most broadly acceptable format, acceptable to almost all respondents. However, it is not their most preferred format, as can be seen in Figure 16.

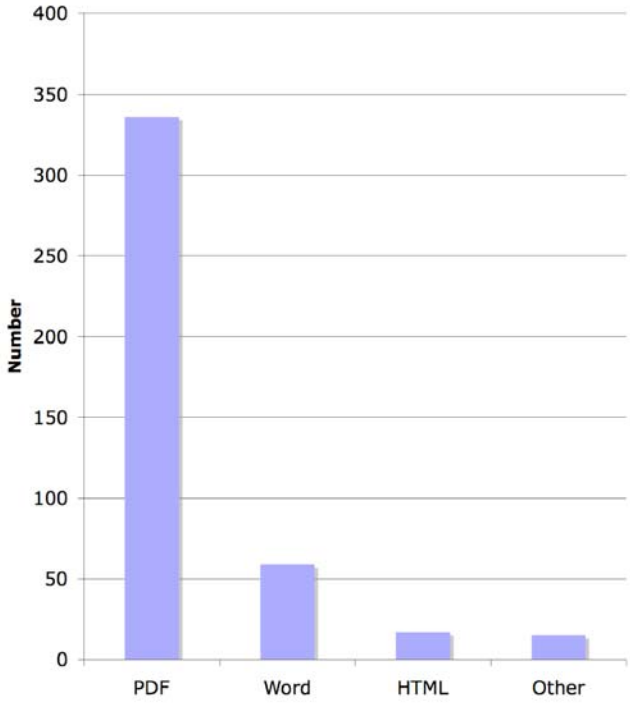


Figure 16: Preferred paper format

Here, most respondents selected PDF as their most preferred format, and no respondents at all selected XML (only four respondents failed to answer this question). There is therefore some distinction between what people are prepared to accept and what people would prefer, and respondents clearly felt strongly that PDF was their favoured choice. While they will tolerate XML, they do not like it, possibly because each publisher implements XML differently and the files are generally not saveable, whereas PDFs create a consistent appearance (as of a printed page) and can be downloaded onto the researcher's computer and not only still be readable but look exactly the same there.

From the point of view of this study, the fact that PDF is so highly preferred is important. It is harder to edit or change PDFs than Word or text documents. Many journals restrict or forbid author self-archiving of the journal PDF.

Knowledge and use of self-archiving

Knowledge of self-archiving

Question 11 gave a brief definition of self-archiving and asked respondents whether they were aware of it. Overall, the majority of people had heard of self-archiving, with nearly 70% of those who answered the question knowing either a little or a lot about it (Figure 17). However, most of these only knew a little, and 30% of respondents had not heard of it at all. This shows an increase in awareness over Swan and Brown's (2005) survey, where only 61% of respondents were aware of self-archiving. As mentioned previously, there might be some self-selection, with people who were not aware of self-archiving being put off from answering or completing the questionnaire; thus these data are likely to be

biased in favour of those who were already aware of self-archiving. However, this is likely to be as true of Swan and Brown (2005) and it is more probable that awareness of self-archiving is increasing over time.

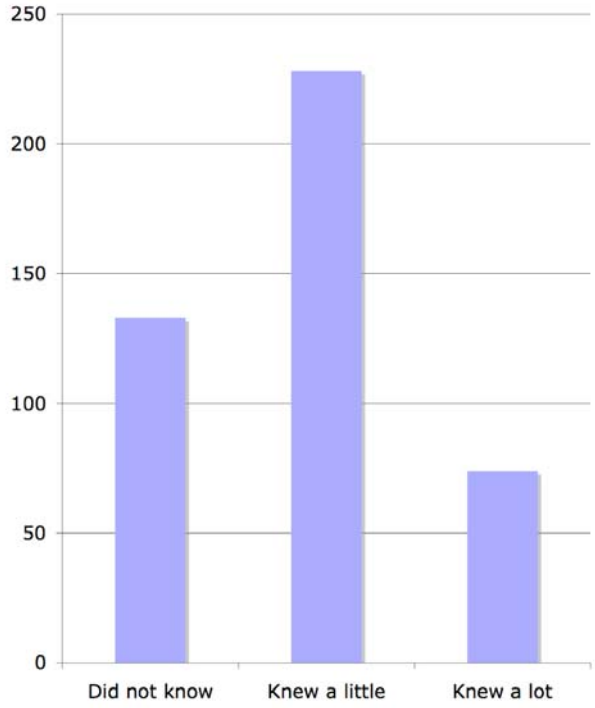


Figure 17: Knowledge of self-archiving

Awareness varied significantly (χ^2=120.7, df=22. p>0.001) between subject fields (see Figure 18). In particular, computer scientists and mathematicians were over twice as likely as expected to know a lot about self-archiving, while those who worked in medicine were very unlikely to know anything about it, with over 60% saying that they did not know about it. This agrees with the findings of Ware (2004a), Ware (2004b) and Andrew (2003), who found little or no medical material in institutional repositories. This is particularly interesting as one of the arguments of open-access advocates is that increased access to medical literature could save lives. It may be that, in those fields where awareness of self-archiving is low, it is because more literature is available in open access journals. Certainly the Directory of Open Access Journals (DOAJ; http://www.doaj.org) lists 210 open access journals in the field of medicine, more than in any other field. Rowlands *et al.* (2004) found that a higher proportion of people knew about open access journals in medical schools than in universities, and Schroter *et al.*'s (2005) survey of *BMJ* authors found that the open access policy of that journal was a major factor in their decision on where to publish. It should be remembered that this survey included GPs who would be more likely to read reviews and summaries than original research, and this might make the awareness of self-archiving appear lower in this field than it actually is. However, even with the GPs excluded from the analysis, 56% of respondents in the field of medicine did not know about self-archiving, a higher proportion than any other field.

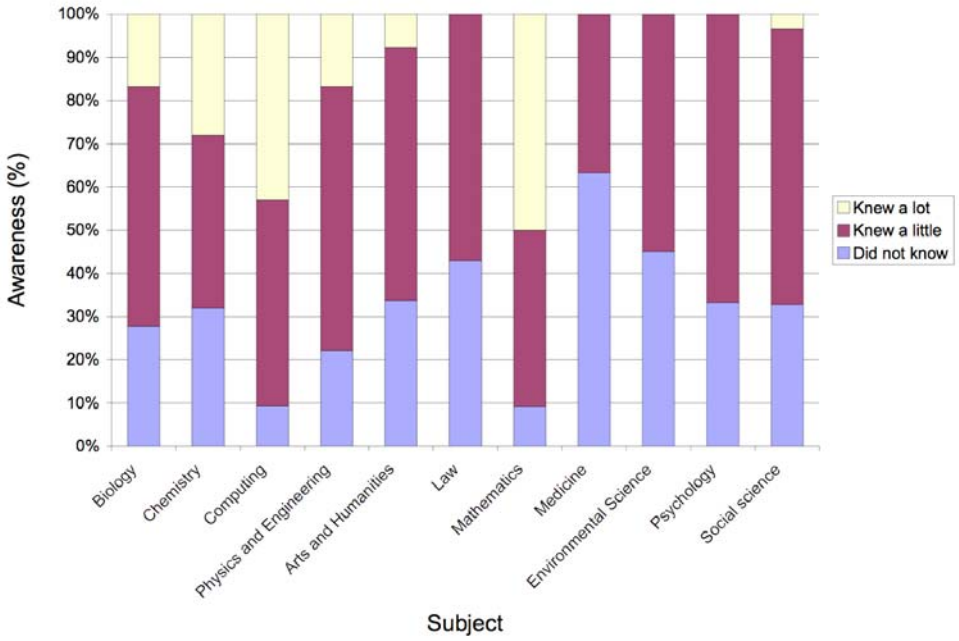


Figure 18: Awareness of self-archiving in different subject fields

There was no correlation between length of time spent in the field and awareness of self-archiving. Minimally active or inactive researchers, as measured by number of papers used, tended to be less likely to be aware of self-archiving than more active researchers (Figure 19), though this was not significant (χ^2=20.2, df=8, NS). That this category also has an increased proportion of respondents who 'knew a lot' implies that this might have more to do with the small sample size than any actual phenomenon; and also relates to the fact that mathematics and medicine, two fields with opposing attitudes to self-archiving, both had a high proportion of respondents using few papers.

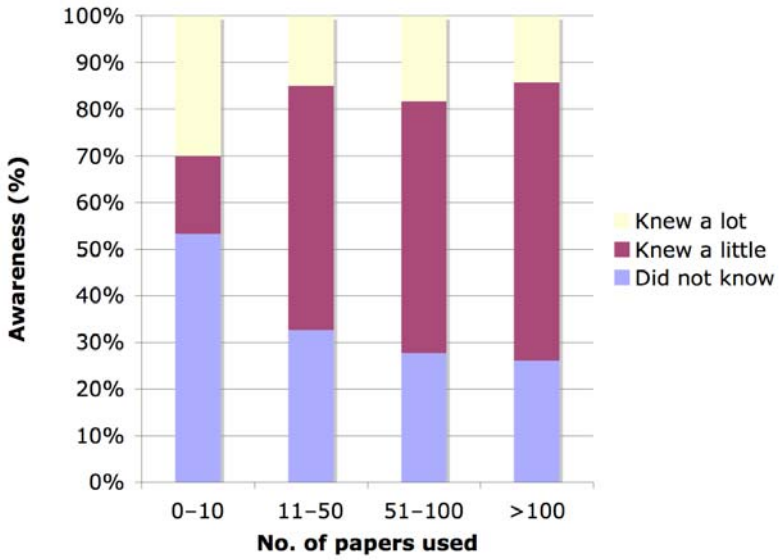


Figure 19: Effect of number of papers used on awareness of self-archiving

Awareness did not seem to be significantly affected by the presence of a repository at the respondent's institution; slightly fewer respondents from institutions with repositories said they did not know about self-archiving, and slightly more said they knew a lot, but this was not statistically significant and could be related to the correlation between subject area and presence of an institutional repository, as discussed earlier.

Respondents who had published a paper were more aware of self-archiving than respondents who had never published a paper (χ^2=10.4, df=2, p<0.01). Awareness was not affected by how recently the last paper had been published (χ^2=12.7, df=8, NS), but was strongly related to the number of papers produced within the past year (χ^2=34.6, df=10, p<0.001), as shown in Figure 21; authors who had not published in the past year had similar levels of awareness to respondents who had never published a paper. Swan and Brown (2005) found a similar phenomenon in their survey, conducted in 2004. The lack of correlation between the number of papers produced and the respondent's subject field (χ^2=78.9, df=55, NS[9]) suggests that this may be an independent effect, although the sample size is too small to be entirely sure that there is no relationship.

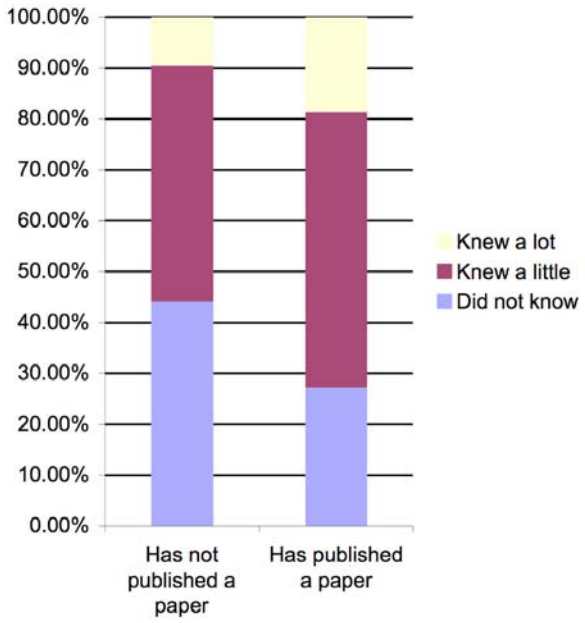


Figure 20: Awareness of self-archiving amongst respondents who had and had not published a paper

[9] Sixty-three per cent of values had expected counts of less than five.

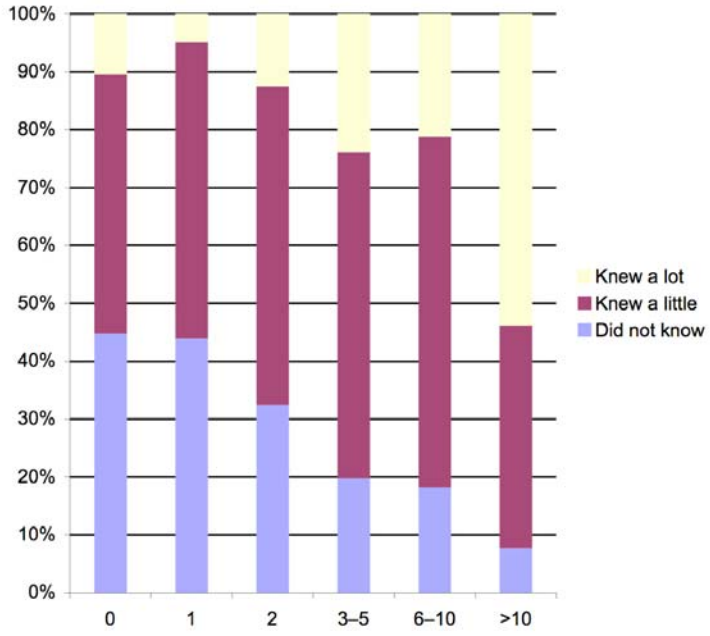


Figure 21: Awareness of self-archiving among authors by papers written in the past year

Understanding of repositories

I followed this with a question designed to establish whether people's general understanding of self-archiving was accurate, and how they valued self-archived material. Figure 22 shows the results.

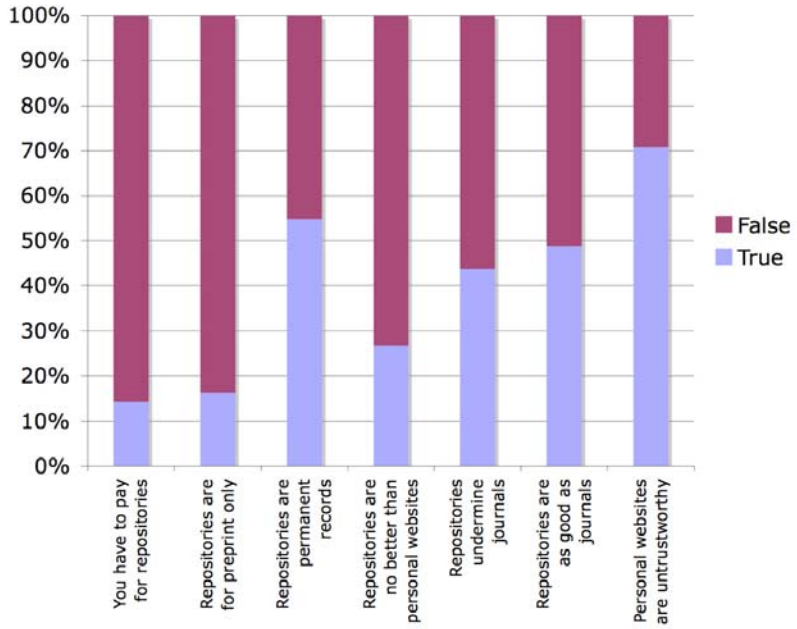


Figure 22: Understanding and valuing of repositories

There was a high proportion of accuracy in the responses to the true/false statements 'You have to pay to place your article in a repository' and 'Repositories are only for pre-prints, not for the final article'. However, only around half the respondents thought that repositories were permanent records of research. Repositories are intended and designed to be permanent, but many people in this sample seemed to feel that they could not trust them to be there in the long term.

The remaining four questions were intended to ascertain the value that people placed on repositories. Over 70% felt that they were better than personal websites, and a similar proportion felt that personal websites were untrustworthy. Fifty per cent felt that repositories were as good as journals; this indicates that, overall, repositories are ranked somewhere between personal websites and actual journals in people's minds. There was a slight trend for more respondents who agreed than disagreed that repositories were no better than personal websites to also believe that personal websites were untrustworthy, but this was not statistically significant (77% as opposed to 68%; χ^2=3.1, df=1, NS).

Awareness of self-archiving modified these opinions somewhat (Figure 23). Respondents who knew about self-archiving were less likely to believe that repositories in general charged authors to deposit articles (χ^2=8.5, df=2, p<0.05) and more likely to think that they were as good as journals (χ^2=26.3, df=2, p<0.001). They also tended to be more likely to believe that repositories were permanent records, although this was not statistically significant (χ^2=5.3, df=2, NS). Interestingly, those who knew a little about repositories were more likely to think that they undermined journal revenue, but this perception was reversed among those who knew a lot about them (χ^2=6.5, df=2, p<0.05). Respondents who were aware of self-archiving also seemed to trust personal websites more (χ^2=6.6, df=2, p<0.05).

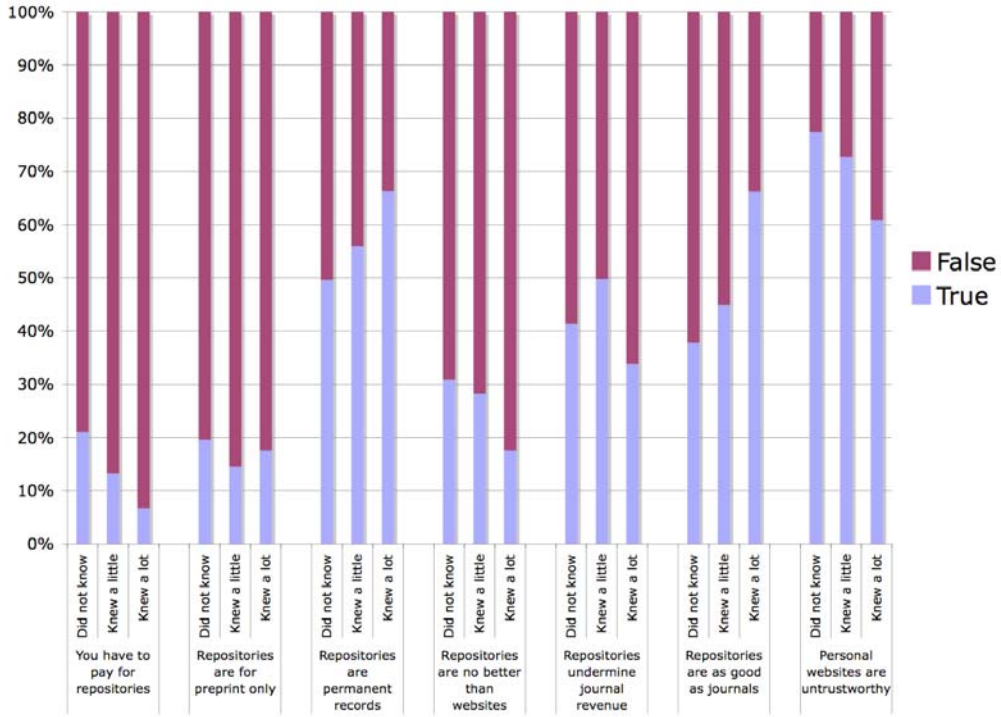


Figure 23: Understanding of repositories affected by knowledge of self-archiving

There were no significant differences between respondents who had been working for different lengths of time in their field, or between less and more active researchers.

Use of self-archived articles

Overall, 310 respondents (71%) had used articles that had been self-archived (Figure 24). While only 43% had used articles from repositories, 65% had used papers obtained from personal websites. Most users of repositories are therefore also downloading articles from websites. The greater use of websites is likely to be for two reasons: first, more material is available on websites than in repositories (Swan and Brown, 2005), and second, websites are more likely than repository articles to be returned on a simple search using the most popular search engine in the survey, Google. In addition, those who are unaware of self-archiving may not know of repositories at all, let alone how to search for articles in them.

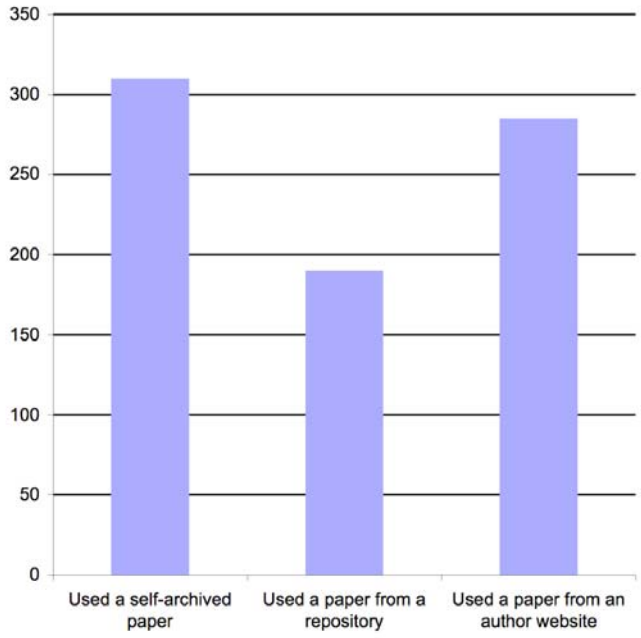


Figure 24: Use of self-archived material

This is borne out by Figure 25, which shows that while knowledge of self-archiving increases use of both repository and website material (χ^2=39.4–102.9, df=2, p<0.001), a large proportion (45%) of those who 'did not know' about self-archiving were already using articles from personal websites. Those who 'knew a lot' about self-archiving were as likely to use articles from repositories as from websites (χ^2=14.2, df=1, p<0.001).

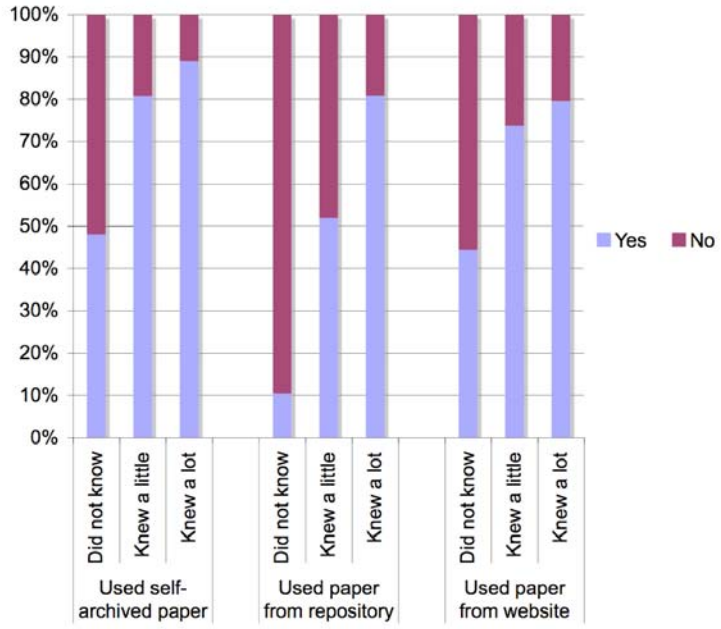


Figure 25: Knowledge of self-archiving and use of self-archived material

As with the knowledge of self-archiving, the use of self-archived materials was affected by the subject area of respondents (χ^2=77.1, df=11, p<0.001). This affected not only the overall number of self-archived articles used but also the proportion of website and repository use (Figure 26): while in mathematics almost as many respondents used repositories as websites, in psychology the proportion using repositories was only a little over half that using websites (47% and 80% respectively). In every field, however, the number of website users was higher than that of repository users. Medicine was a particular outlier, with only 35% of respondents in that field having used any self-archived articles, whereas all the other fields had at least 60%.

Some respondents pointed out in their comments that one reason for the different behaviour in different fields might be the differing importance of peer review and speed of publication:

> If I was looking for scientific research I would be more worried about the reliability of the paper and whether it came from a trustworthy source. But since my research is in philosophy, a bad paper is simply one that's reasoned out badly, or has an invalid argument, and I should be able to tell whether this is the case just by reading it!

> The prevalence of self-archiving seems to depend strongly on the particular field. In experimental neuroscience, self-archiving is more rare than in, say, mathematics or elementary particle physics, perhaps because the correctness and scientific quality of the neuroscience papers is less evident to someone from a different subfield. Thus, peer review and the prestige of journals … is quite important in neuroscience.

> Currently I am working in a very contentious area and it is vital to get stuff out quickly. When I work in less contentious areas I am less concerned to get stuff on my website.

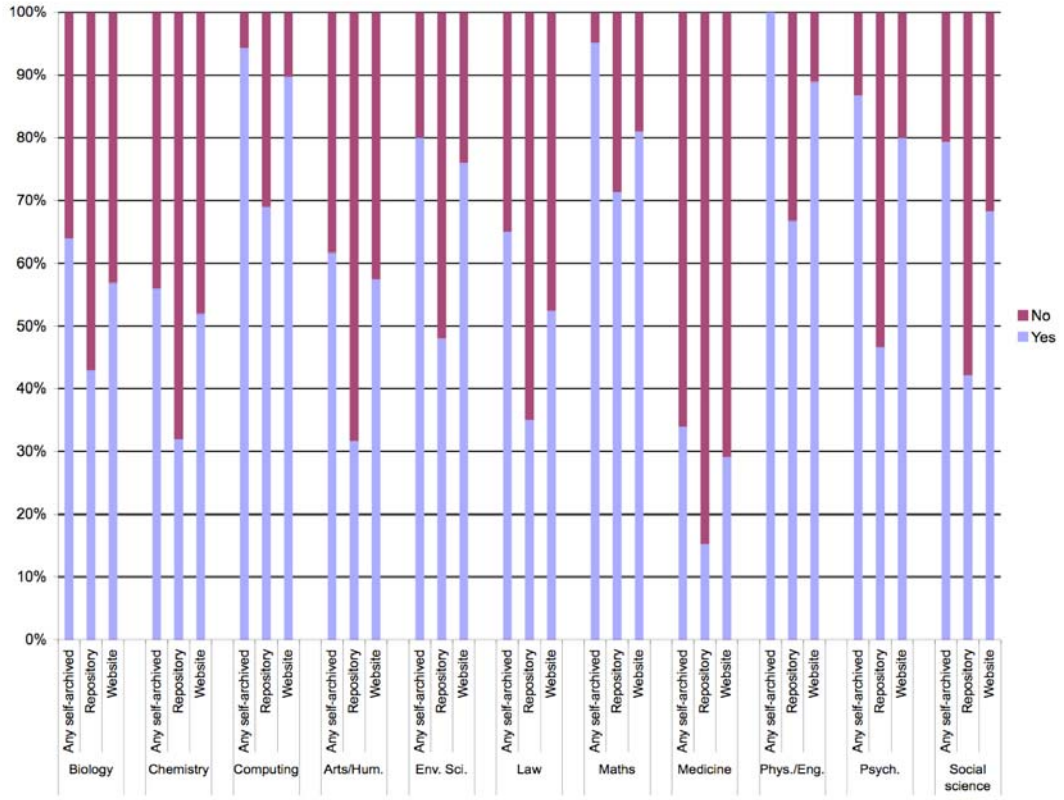


Figure 26: Effect of subject field on use of self-archived articles

Use of self-archived articles was also affected by role (Figure 28; χ^2=30.0, df=10, p<0.001) and number of years in the field (Figure 28; χ^2=12.7, df=4, p<0.05). With the senior lecturers, readers and GPs discounted (all had only a very small sample, with six respondents or fewer each, although reasons for GPs' lack of use of self-archived articles are discussed elsewhere), more junior researchers do indeed appear to be more likely to have used self-archived articles than more senior ones. There was also a significant relationship with productivity in terms of number of papers produced in the previous year (Figure 29), with more productive authors being more likely to have used self-archived articles (χ^2=15.9, df=5, p<0.01). As there is no significant relationship between use of self-archived material and number of articles *used* (χ^2=8.3, df=4, NS), I speculate that authors are being made aware of self-archiving, perhaps by advocates of their institutional repository, as an option for publishing their papers, rather than researchers in general being made aware of self-archiving as an option for finding papers. Swan and Brown (2005) found that more prolific authors were more likely to self-archive; this study shows that they are also more likely to use other people's self-archived articles.

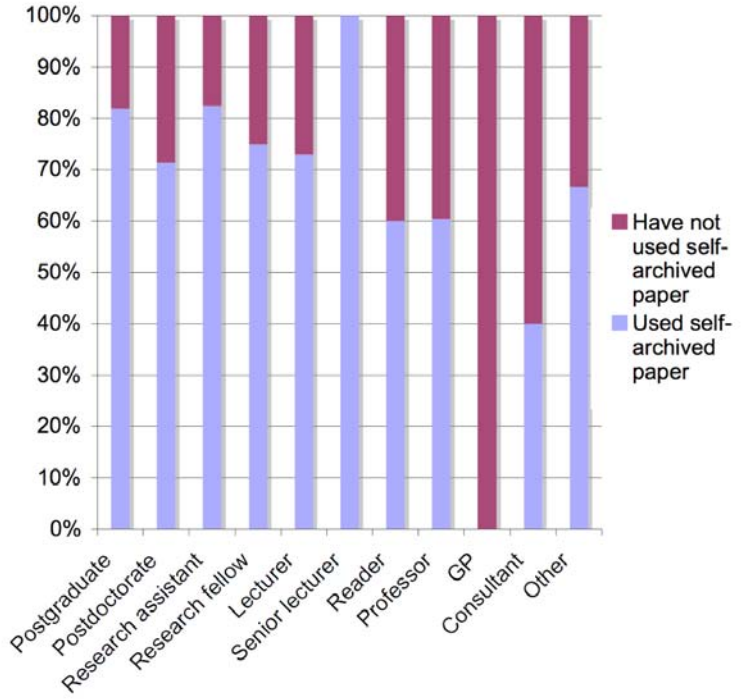


Figure 27: Use of self-archived material by role of respondents

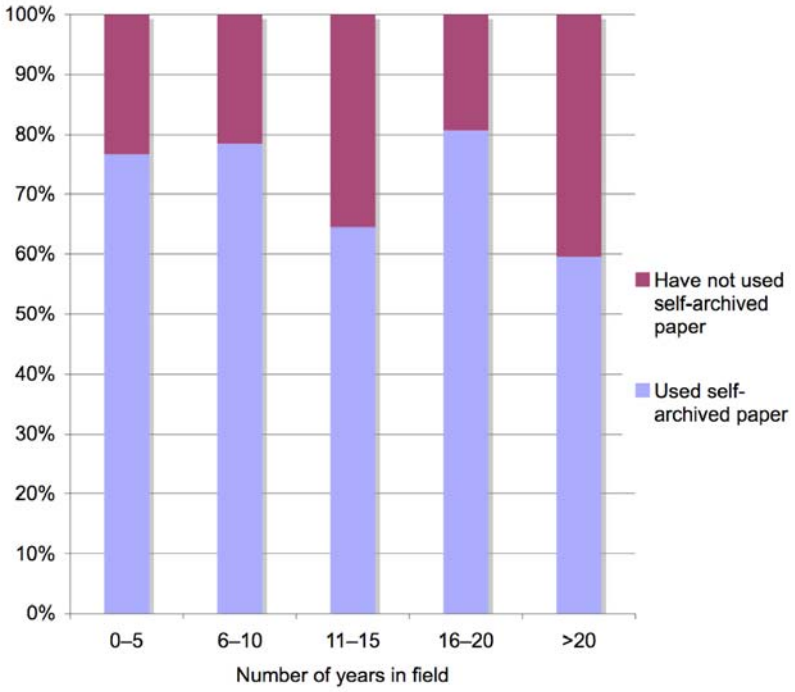


Figure 28: Number of years in field and use of self-archived material

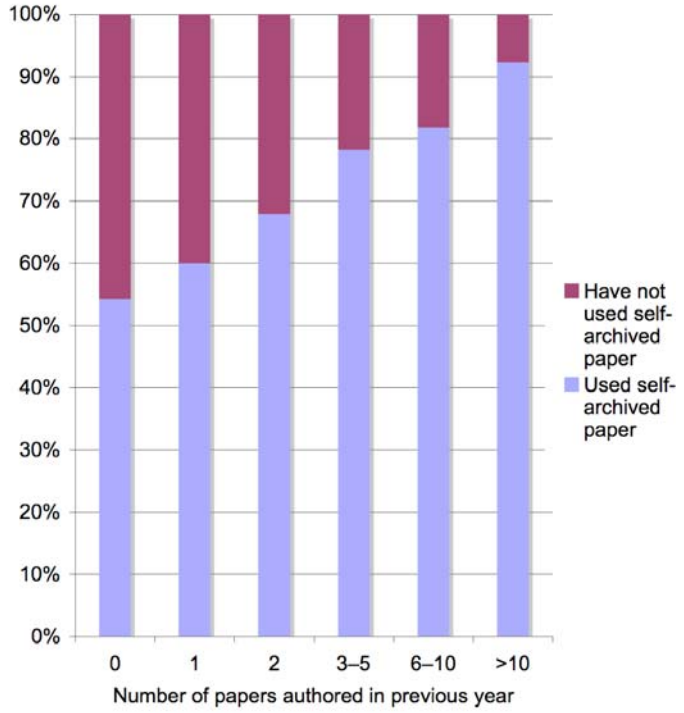


Figure 29: Use of self-archived material by number of papers produced in previous year

Another factor affecting the use of self-archived materials was the choice of search facilities. Respondents who used search facilities that included self-archived works were more likely to have used self-archived papers (Figure 30). This is, on the face of it, obvious, as they are more likely to come across such material. Interestingly, a large proportion (45%) of respondents who use facilities that only return publisher-based material have used a paper taken from an author's website: they may be using other, undeclared search facilities to find these, or, more likely, they are accessing them through direct links from other material or personal communication with the authors or others.

It is worth examining this in more detail. It may be that the choice of search facility is influenced by the knowledge of self-archiving, such that people who know more about self-archiving deliberately seek out sources of self-archived material. Figure 31 suggests that this is at least part of the explanation, with respondents who 'knew a lot' being more likely to use a range of search facilities that included both publisher-based and self-archived material. However, the choice of search facilities for respondents who only 'knew a little' was similar to that of those who 'did not know' anything: less than half of respondents in these two groups were using a set of facilities that would allow them to find both varieties of material.

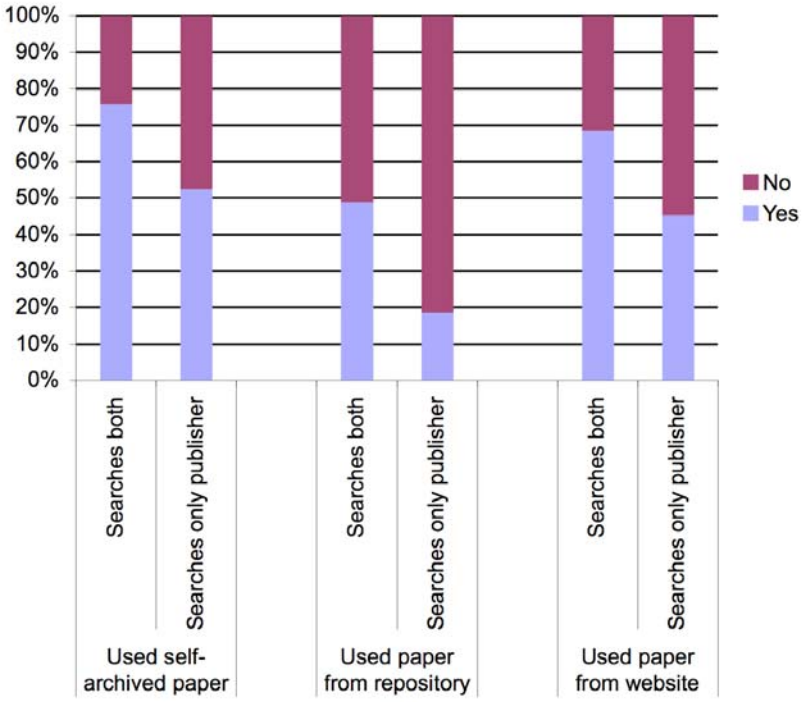


Figure 30: Search facilities and use of self-archived material

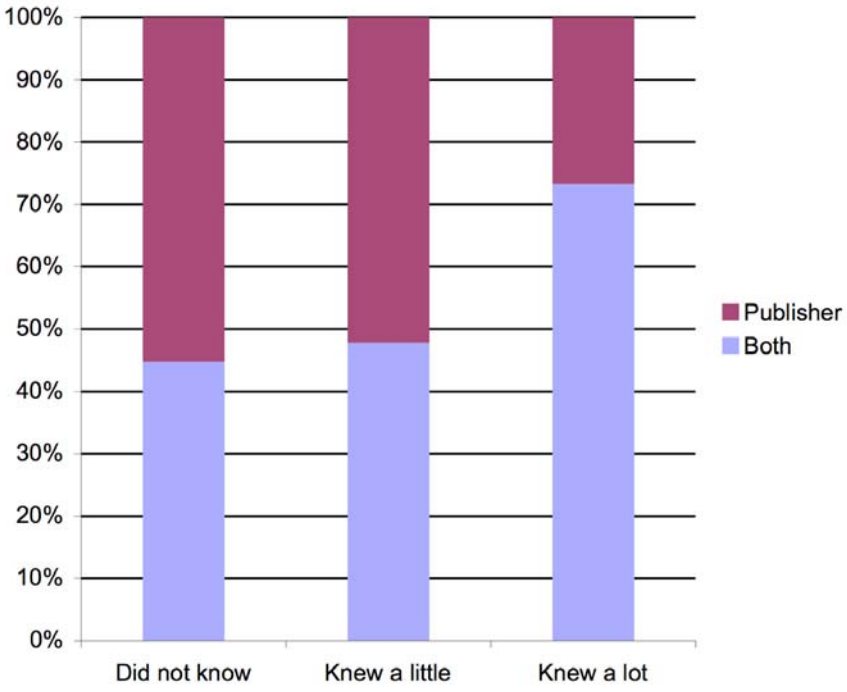


Figure 31: Knowledge of self-archiving and choice of search facility

In many cases, the individual search facilities chosen did not affect respondents' use of self-archived material (and some had too few users in this survey to draw any firm conclusions). However, for the

most popular search facilities, users of all but Medline and PubMed showed greater than average use of self-archived material (Figure 32), while for Medline and Pubmed, users were less likely than average to use self-archived material, while non-users were more likely than average. It is probable that the users of Pubmed and Medline, many of whom are in the field of medicine, are bringing down the average because of their subject field's bias against self-archiving.

While the individual search facilities did not affect whether or not respondents had used self-archived material, Figure 33 shows that the number of different search facilities used had a definite effect. Overall, the number of search facilities chosen peaked broadly between 2 and 7, but this peak is actually made up of two separate peaks for the users and non-users of self-archived material: non-users peak at two facilities whereas the users peak at seven (χ^2=34.3, df=17, p<0.005). The users are therefore searching in a greater variety of ways and might thus be more likely to come across self-archived material. This does not correlate with knowledge of self-archiving (χ^2=24.6, df=17, NS), suggesting that this might be an independent phenomenon. (The sample size is too small to tell whether there is any correlation with the subject field, but as these relate so strongly this is unlikely.)

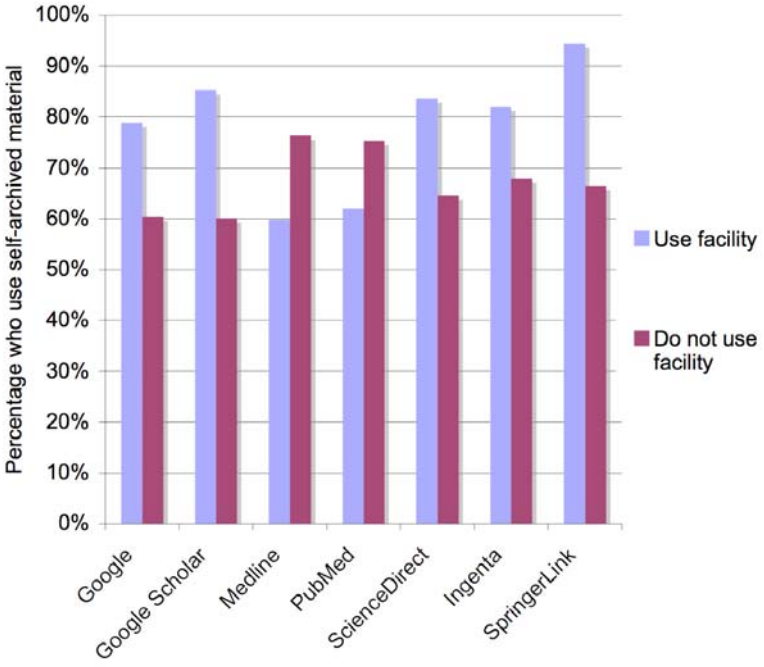


Figure 32: Use of self-archived material among users of the most popular search facilities (overall average = 71%)

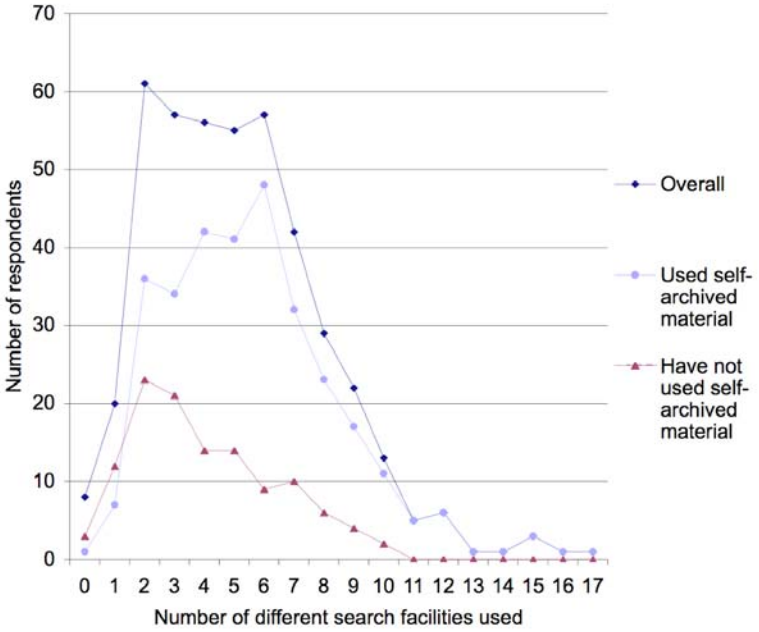


Figure 33: Number of respondents using multiple search facilities

Use of repositories

Question 13 asked whether the respondent's institution had a repository, and the results are shown in Figure 34.

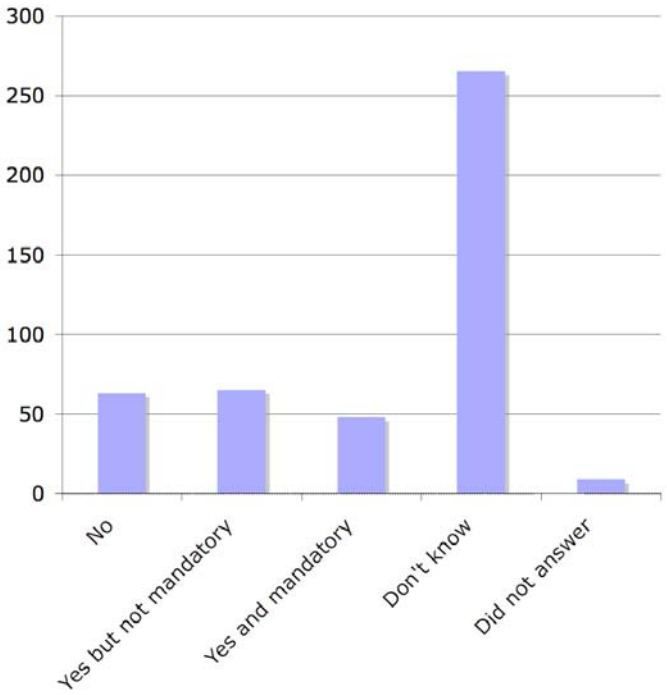


Figure 34: Perceived repository status

Over 60% of respondents who answered the question did not know whether their institution had a repository. Figure 35 shows that even among those who did know a little about self-archiving, just over 55% said that they did not know whether their institution had a repository or not. Widespread awareness of the existence of an institutional repository only occurred among those who knew a lot about self-archiving.

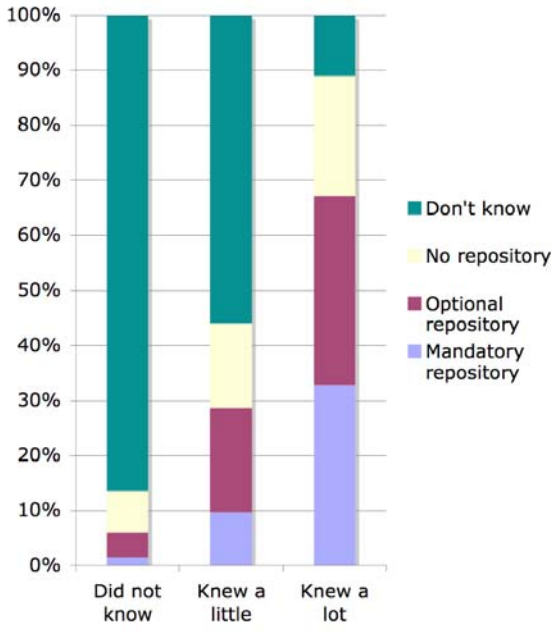


Figure 35: Knowledge of institutional repositories by knowledge of self-archiving

The overall level of knowledge, however, was only slightly higher in institutions with repositories than in institutions without, as shown in Figure 36. Among those who did not choose 'Don't know' in response to this question, knowledge of whether their institution had a repository was in general fairly accurate.

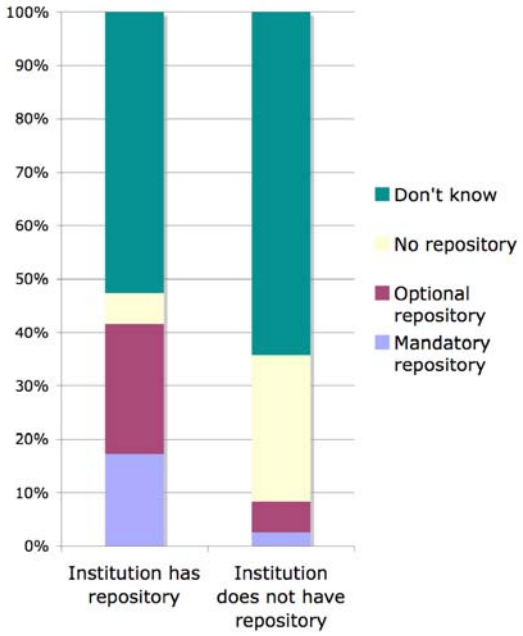


Figure 36: Perceived and actual institutional repositories

However, 8% of respondents said that their institution had a repository when in fact it did not. There are several possible reasons for this:

- Respondents might be considering repositories hosted by some institution other than their main affiliated institution given in the answer to Question 1.
- The institutions might have repositories that are not listed on ROAR or OpenDOAR, or are restricted to members of the institution and thus inaccessible to external users.
- Respondents might be aware that their institution is developing a repository even though it is not yet fully operational, and are answering the question on this basis.
- Respondents might be confused about the definition of a repository or are otherwise answering incorrectly.

Forty-eight respondents (11%) stated that they were obliged to place their articles in their institution's repository. According to the information I gathered from the Registry of Open Access Repository Material Archiving Policies (ROARMAP; http://www.eprints.org/openaccess/policysignup/), ROAR, OpenDOAR, institutional repository websites and enquiries to repository managers (see Appendix 2), only one repository in this study (the departmental repository for the Department of Electronics and Computer Science at the University of Southampton) currently has a policy of mandatory deposit, although some others are planning to introduce one soon. (It should be noted, however, that I could not obtain full information about some of the repositories.) Twenty of these respondents came from this department, leaving 28 respondents (6%) who believed that they were obliged to self-archive when they were not. This could be for several reasons:

- Respondents might be considering policies of their funders that mandate deposit in a repository as a condition of funding. (However, only eight respondents claimed that they were obliged to self-archive by their funders, and there was no significant difference between their answer to this question and the answers of other respondents; χ^2=0.42, df=2, NS.)

- Institutions might have policies of mandatory deposit that are not listed in ROAR, ROARMAP or OpenDOAR and are not shown on the repository website.

- Respondents might be aware that their institution is moving to a policy of mandatory deposit and are answering the question on this basis.

- Respondents might be confused about their institution's policy, particularly if the repository has been promoted aggressively.

Figure 37 shows that 167 respondents (just under 40% of those who answered the question) had used articles from institutional repositories.

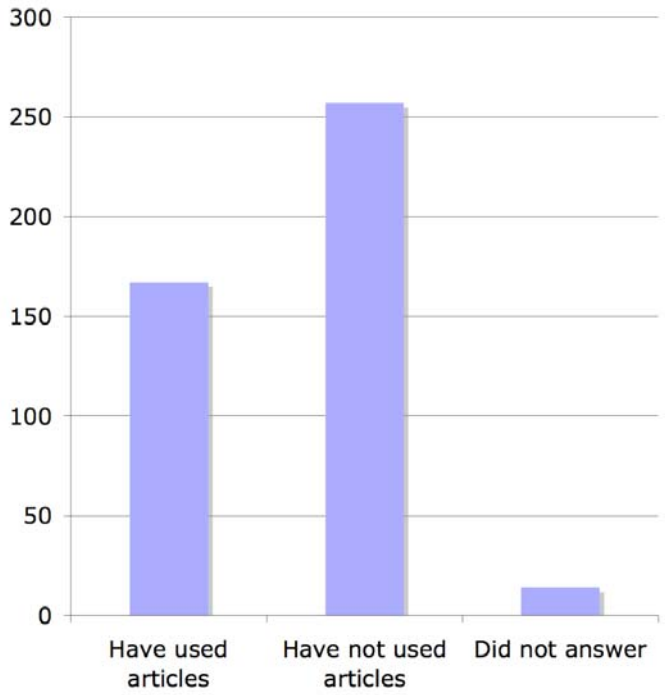


Figure 37: Use of articles from institutional repositories

Unsurprisingly, this correlated strongly with people's awareness of self-archiving, with respondents who knew about it being much more likely (χ^2=86.6, df=2, p<0.001) to have used articles from an institutional repository (Figure 38). Moreover, those who believed that their institution had a repository were more likely to have used articles from an institutional repository than those who believed that it did not, and people who did not know if their institution had a repository were less likely again (Figure 39; χ^2=55.3, df=3, p<0.001). This is probably related to their knowledge of self-archiving, and it is likely that many of those respondents who know that their institution has a repository have used it. However, the actual

presence of a repository at their institution did not have a significant effect on respondents' answer to this question (χ^2=2.3, df=1, NS), indicating that at least some researchers are looking in other institutions' repositories for information.

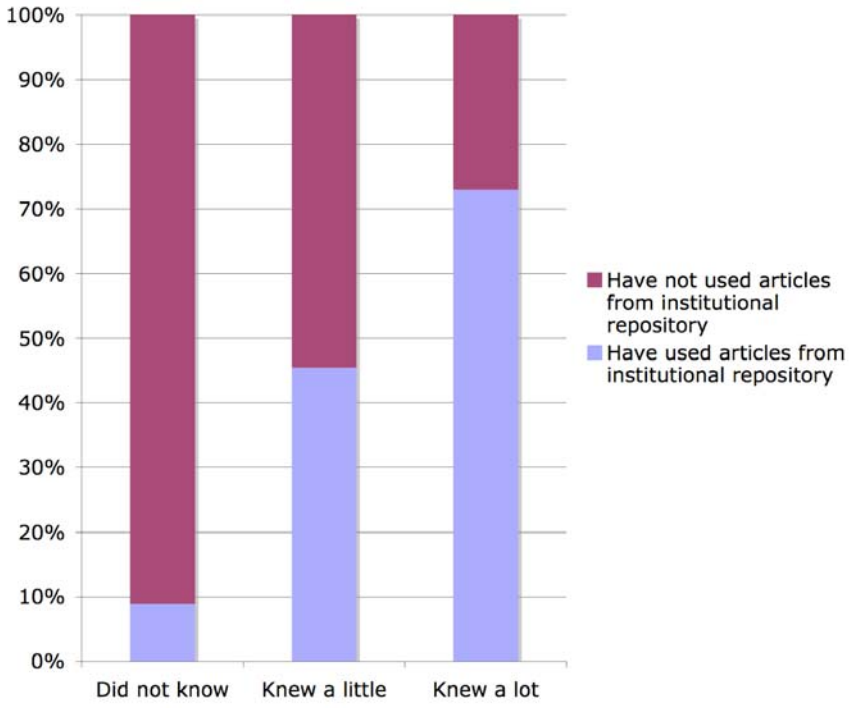


Figure 38: Knowledge of self-archiving affects use of institutional repositories

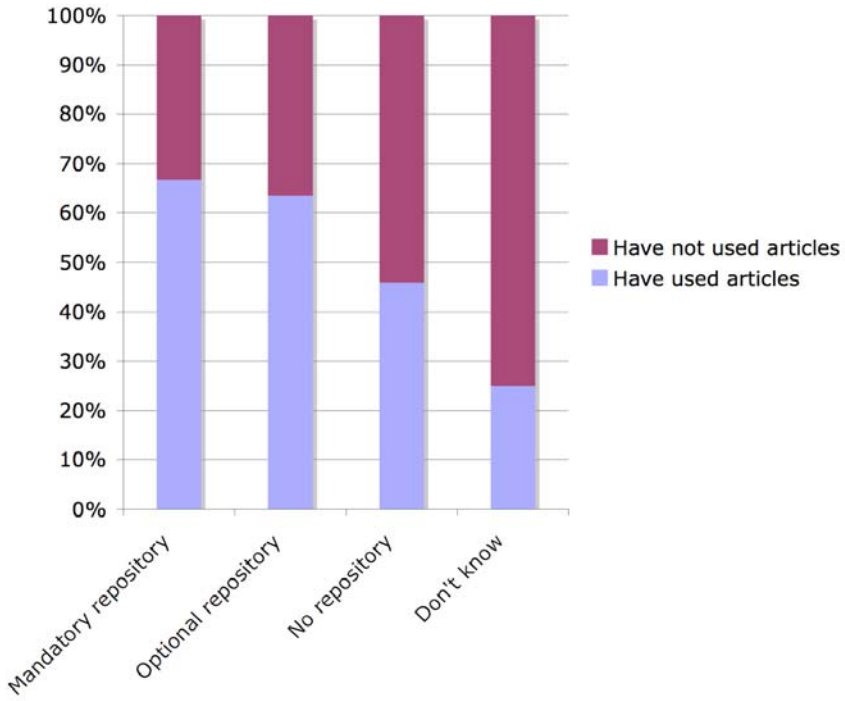


Figure 39: Use of institutional repositories and perception of institutional repository status

Use of subject-based repositories

Figure 40 shows the level of awareness and use of subject-based repositories.

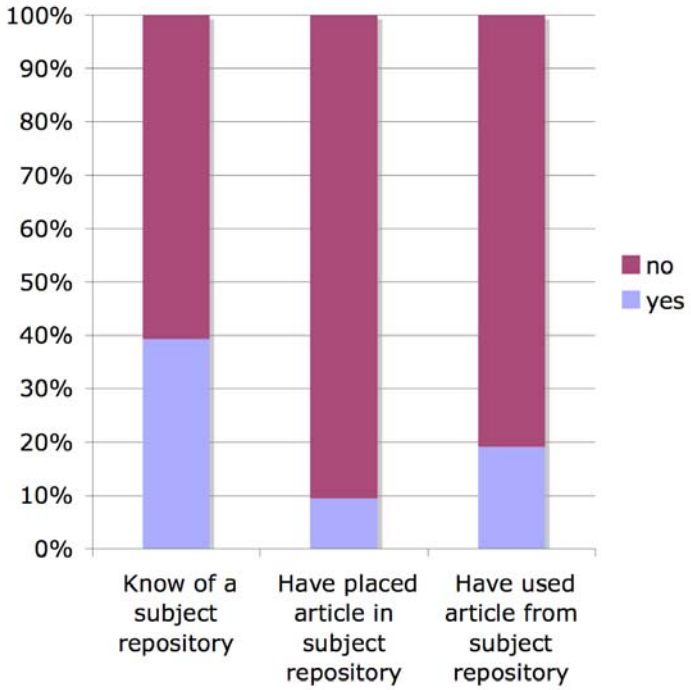


Figure 40: Awareness and use of subject-based repositories

Fewer than 20% of respondents had used articles from a subject-based repository (around half the number that had used articles from an institutional repository; almost all of those who had used subject-based repositories had also used institutional repositories). Figure 41 shows that respondents in the fields of mathematics and computing were particularly likely to use subject-based repositories, while chemists and psychologists were particularly unlikely to (χ^2=42.7, df=10, p<0.001). This is likely to be related to the existence or lack of suitable repositories in each field, or to different amounts of awareness due to varying publicity for and advocacy of the different subject repositories. This would also help to explain the different levels of awareness of self-archiving in different fields (see Figure 18).

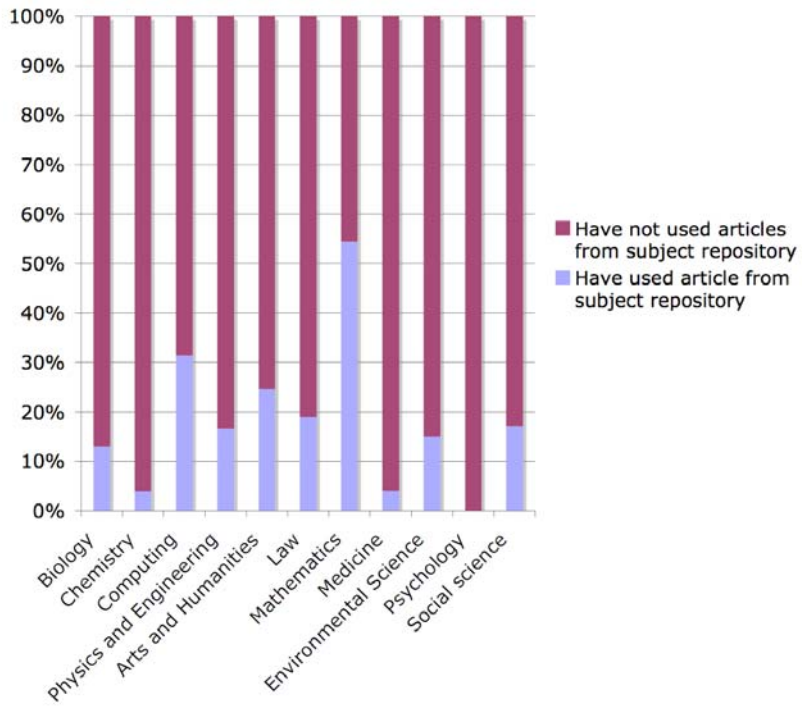


Figure 41: Use of subject-based repositories by subject

Even though medicine has already been shown to be a field with a particular lack of awareness of self-archiving, the low use of subject repositories in medicine is perhaps surprising: one of the major subject repositories, PubMed Central, is based in this field. This is easily accessed through Pubmed, which is used by 70% of respondents in the field of medicine. Pubmed or Medline is the preferred source of articles for over 40% of respondents in the field of medicine.[10] It is possible that the respondents did not consider this to be a subject repository, as the information on the website refers to it as an 'archive' rather than a 'repository' or a 'database'; it may also be possible that respondents are unaware of PubMed Central's existence, perhaps because they access Medline through a third-party leasing arrangement which does not provide links to PubMed Central.

[10] Pubmed contents are included in almost all Medline leases (National Library of Medicine, 2004).

There was no statistically significant relationship between use of subject-based repositories and length of time in the field or number of papers used per year.

Reasons for use of self-archived article

Question 16 asked, firstly, whether respondents had ever placed a copy of their article on their own website. Of those who answered the question, 23% had. This rises to 28% when those who had never published a paper were excluded. Respondents were more likely (χ^2=40.2, df=2, p<0.001) to have done so if they were aware of self-archiving, but those who knew most were less likely to have done so than those who knew only a little. This might be because those who are more aware of self-archiving are eschewing personal websites in favour of uploading their articles to formal repositories.

The remainder of the question asked whether respondents had used articles from repositories and authors' websites, and under what circumstances they had done so. The results are shown in Figure 42.

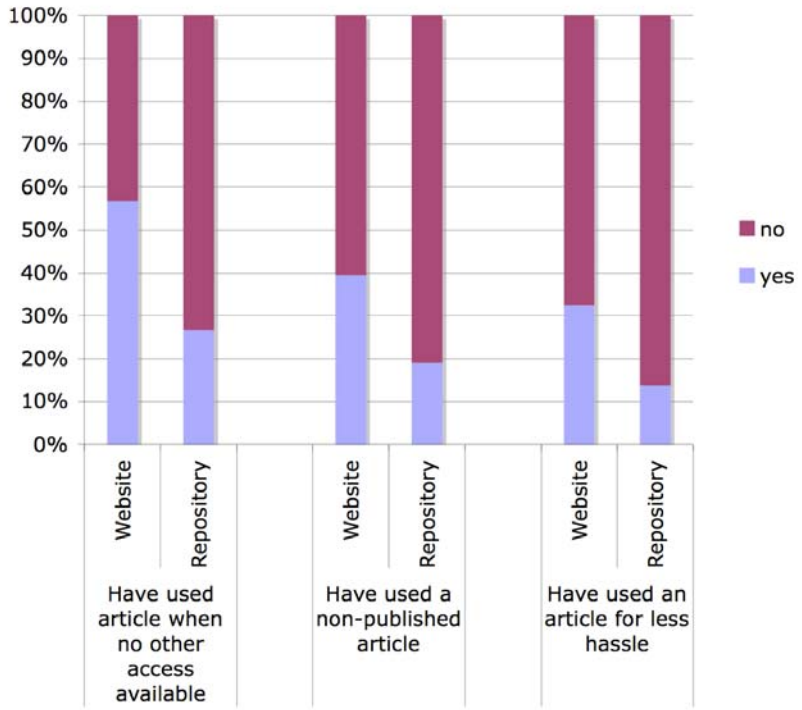


Figure 42: Use of articles from repositories and authors' websites

In all cases, respondents were around twice as likely to download articles from personal websites than from repositories (χ^2=75.8–101.2, df=2, p<0.001). Lack of access to the journal is the strongest motivator for downloading an article, with over 50% of respondents having downloaded an article from a personal website for this reason, while 27% of respondents had used a repository for the same purpose. Just under 40% of respondents had downloaded articles from websites that had not been published elsewhere; comments from some respondents indicated that these were likely to be reports, conference presentations and other non-journal material, as well as preprints of articles in press.

Only 14% of respondents had downloaded an article from a repository even though they had access to the journal that carried it. This increased to 32% for websites. Possible scenarios where this would occur include: having access only to the print version and desiring the electronic version; having institutional access but wanting to access it from a non-institutional network; being unable to remember or unwilling to bother with the password for a restricted site; and using search facilities that listed both versions and choosing the first or most convenient link presented. In any event, this suggests that at least some respondents are prepared to use self-archived articles for the sake of an easier life, and this implies that these respondents believe that the two versions are equivalent in value.

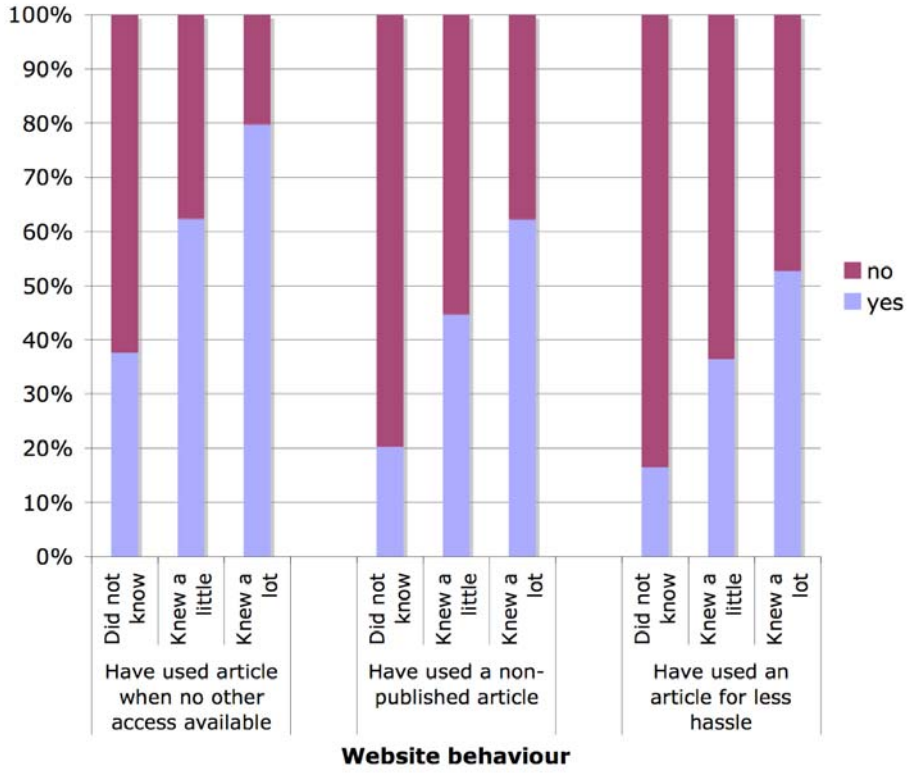
As Figure 43 shows, awareness of self-archiving has a large impact on all forms of use of non-publisher sources (χ^2=30.4–71.4, df=2, $p<0.001$). The effect was stronger with repositories than with personal websites (because people are finding articles on personal websites even if they do not know about self-archiving), and those who knew a lot about self-archiving were more likely to use non-publisher sources than those who only knew a little. Nevertheless, these respondents were still more likely to download articles from websites than repositories. This may be because there is not yet enough material in repositories, because the respondents cannot easily search for articles in repositories, or because the repositories (and search facilities to find material in them) are recent and the occasion to search for a paper outwith publisher sources has not yet arisen since they were established. Some respondents commented that repository sources were difficult to find, or that they might not have known that they were using them:

> I get a lot of papers through general internet searches, so I have never knowingly used a repository, but perhaps it's not always going to be obvious to me where papers are being hosted when I find them.

> If a Google search found an article in a free archive I would then know about the archive, but so far Google has always returned self-archived papers [from personal websites] or papers in repositories that charge for access.

I did not ask how often or how recently respondents had done any of these; these might be questions to ask in future research.

Respondents who were aware of self-archiving were also much more likely to use articles from author websites or repositories because it was easier for them: over 40% of respondents who 'knew a lot' about self-archiving had used an article from a repository even though they had access to the subscriber version, because it was less hassle. In other words, these people are using the repository versions, at least occasionally, in preference to the subscriber versions.



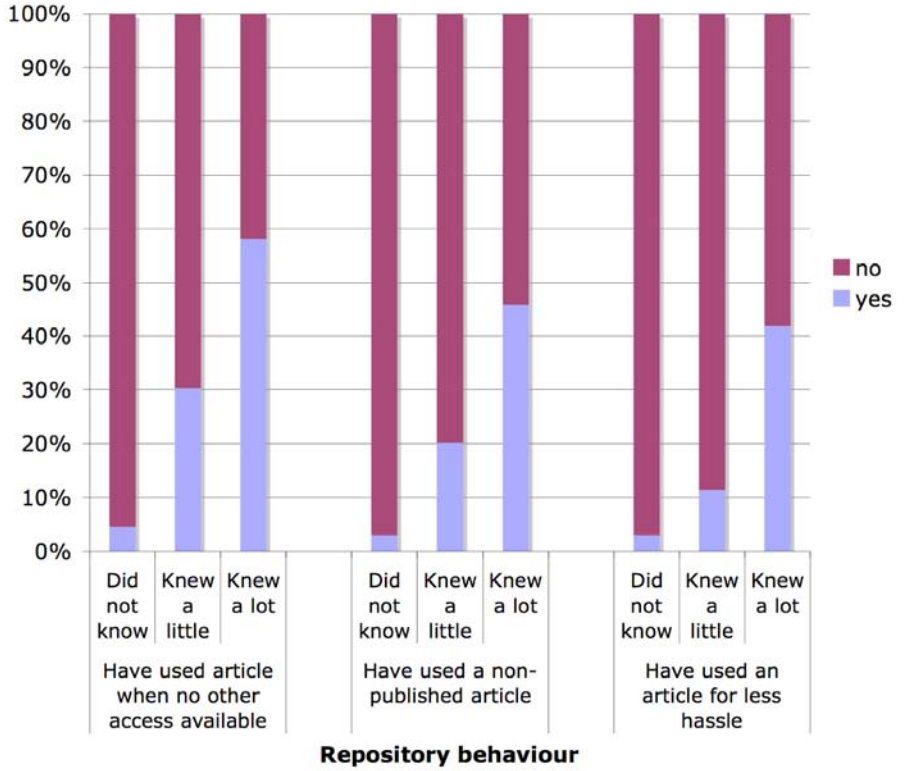


Figure 43: Awareness of self-archiving affects accessing from non-publisher sources

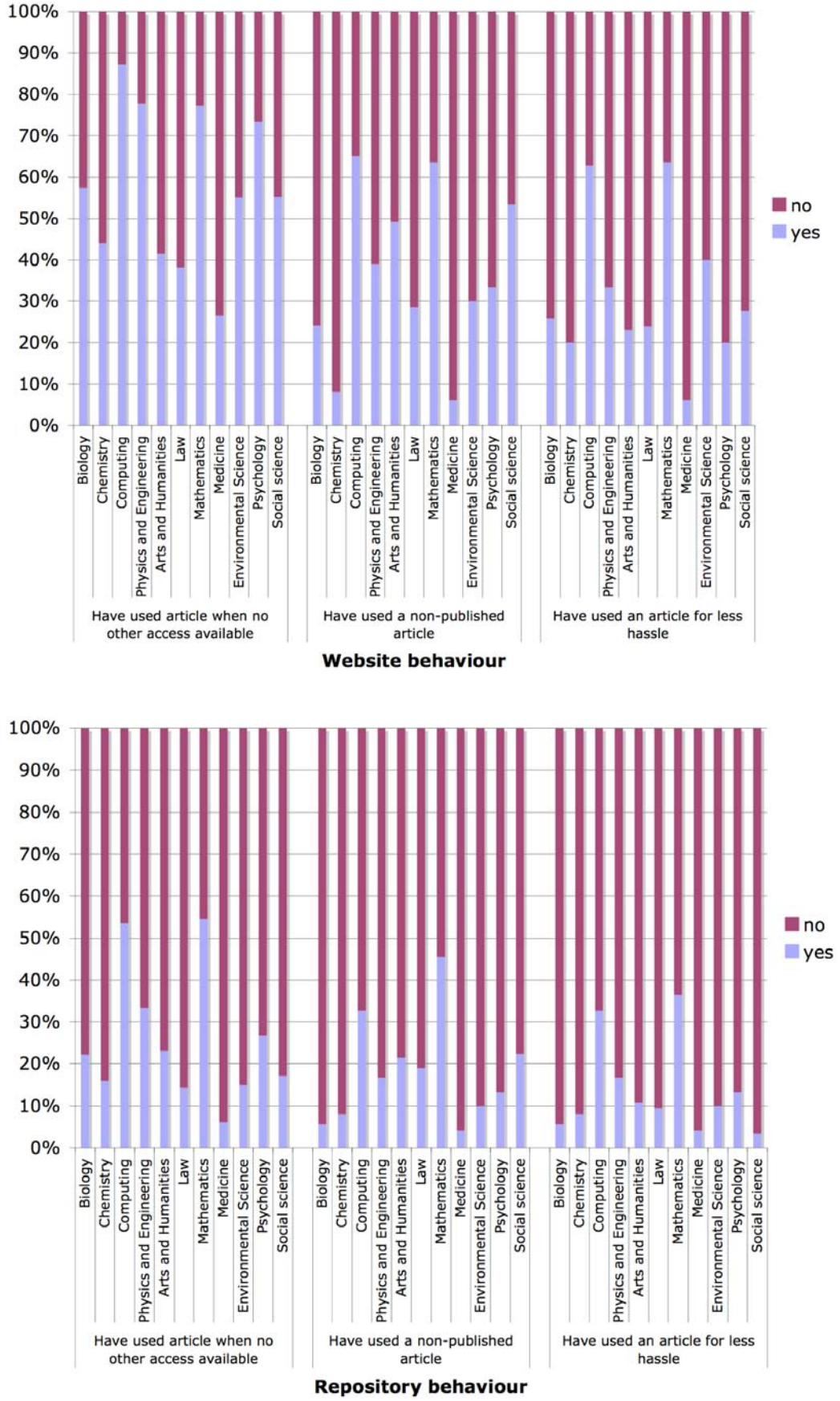


Figure 44: Subject influence on use of articles from non-publisher sources

Figure 44 shows that, as expected, use of repositories for all reasons is more likely (χ^2=48.1–76.2, df=10, p<0.001) amongst those subjects where there is a high knowledge of self-archiving (Computing and Mathematics). However, in website use, they are joined by other fields, notably Physics and Engineering and Psychology. In particular, use of author websites to obtain articles published in journals that the respondent does not have access to is common in almost all disciplines; only in Medicine does it fall below 30%. Also notable is the increased use of author websites to access unpublished articles in Arts and Humanities and Social Science. It is possible that these disciplines make more use of non-journal-based articles, as described above, than other fields.

Only 9% of respondents had used more than one version of the same article.

Views of self-archiving

Question 17 presented a list of statements with the option to agree or disagree with each one; Figure 45 shows the basic results.

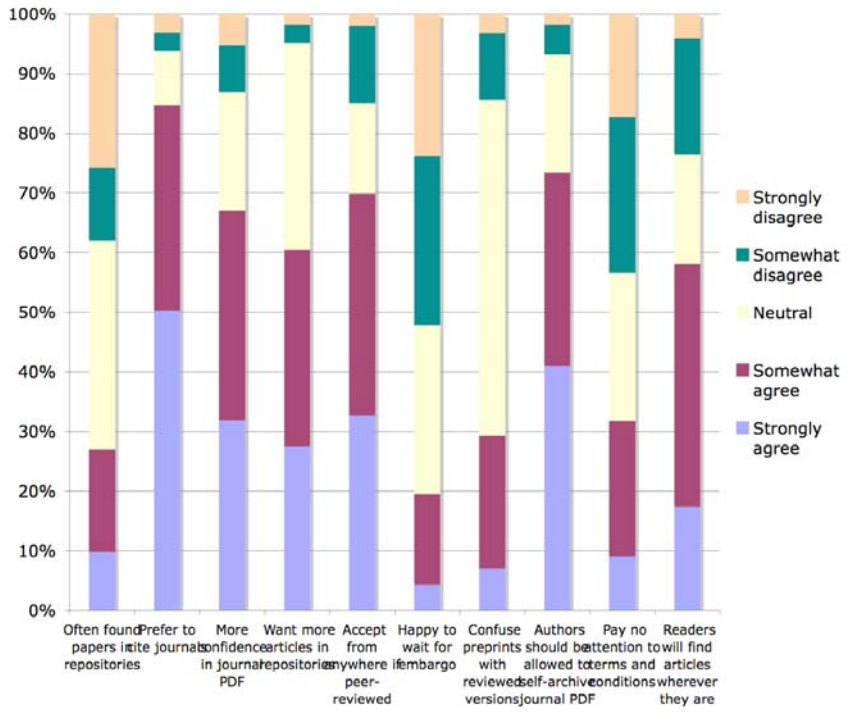


Figure 45: Attitudes to factors relevant to self-archiving

Only 27% of respondents agreed or strongly agreed that they had often found papers in repositories. A greater number, 38%, disagreed or strongly disagreed with the same statement.

There was a very strong preference in favour of the journal article, both in terms of referencing (84% of respondents agreed or strongly agreed with the statement 'I would prefer to cite an article I'd found in a journal rather than one I'd found in a repository or on a website', indicating that they would rather give an 'official' citation than a reference to a website or repository ID) and in terms of the format of the article, where 67% agreed or strongly agreed with the statement 'I have more confidence in the journal PDF

than in the author's own version'. However, despite this, just over 60% wanted more articles to be available in repositories and 70% were happy to accept articles from anywhere as long as they had been peer-reviewed. So while a great deal of authority inheres to the official journal citation and the official journal PDF version of a paper, readers would nevertheless like the opportunity to access papers through other sources. Very few people (under 20%) were happy to wait for the standard embargo term of 12 months for an article to become available in repositories, and 74% believed authors should be allowed to use the journal PDF version for self-archiving. Taken together, this suggests that people are prepared to use repositories to find articles, but they want the journal PDF version and they want it immediately.

While a third of respondents agreed or strongly agreed with the statement 'I don't pay much attention to the publishers' terms and conditions when I publish a paper', 43% disagreed or strongly disagreed. Some respondents suggested that the problem was not that they did not pay attention, but that the conditions themselves were unclear:

> For all sorts of reasons, the rules on copyright (with respect to journal/book publications and indeed use of other texts within these) absolutely need to be clear and consistent. My experience with copyright … is that this is absolutely not the case.

> Some allows reprint after a year or so. Some authorises the author to reproduce the work for personal use. Does that mean it is ok for public accessible self-archiving? No one knows…

As expected, awareness of self-archiving affects the responses to almost all these statements (Figure 46). An exception was the response to the statement 'I would prefer to cite an article I'd found in a journal rather than one I'd found in a repository or on a website' (χ^2=9.5, df=8, NS). Eighty per cent or more of respondents in all three categories agreed or agreed strongly with this statement. This suggests that even those respondents who use self-archived articles regularly (over 60% of respondents who 'knew a lot' about self-archiving agreed or strongly agreed with the statement 'I have often found articles relevant to my work in repositories', as opposed to only 8% of those who 'did not know'; χ^2=102.5, df=8, p<0.001) want to be able to give an official journal citation in their own papers. The journal publisher's input is therefore still highly valued, even amongst those respondents who do not feel that the journal PDF itself adds significant value (less than half of those respondents who 'knew a lot' agreed or strongly agreed with the statement 'I have more confidence in the journal PDF than in the author's own version', as opposed to 74% of those who 'did not know'; χ^2=26.3, df=8, p=0.001).

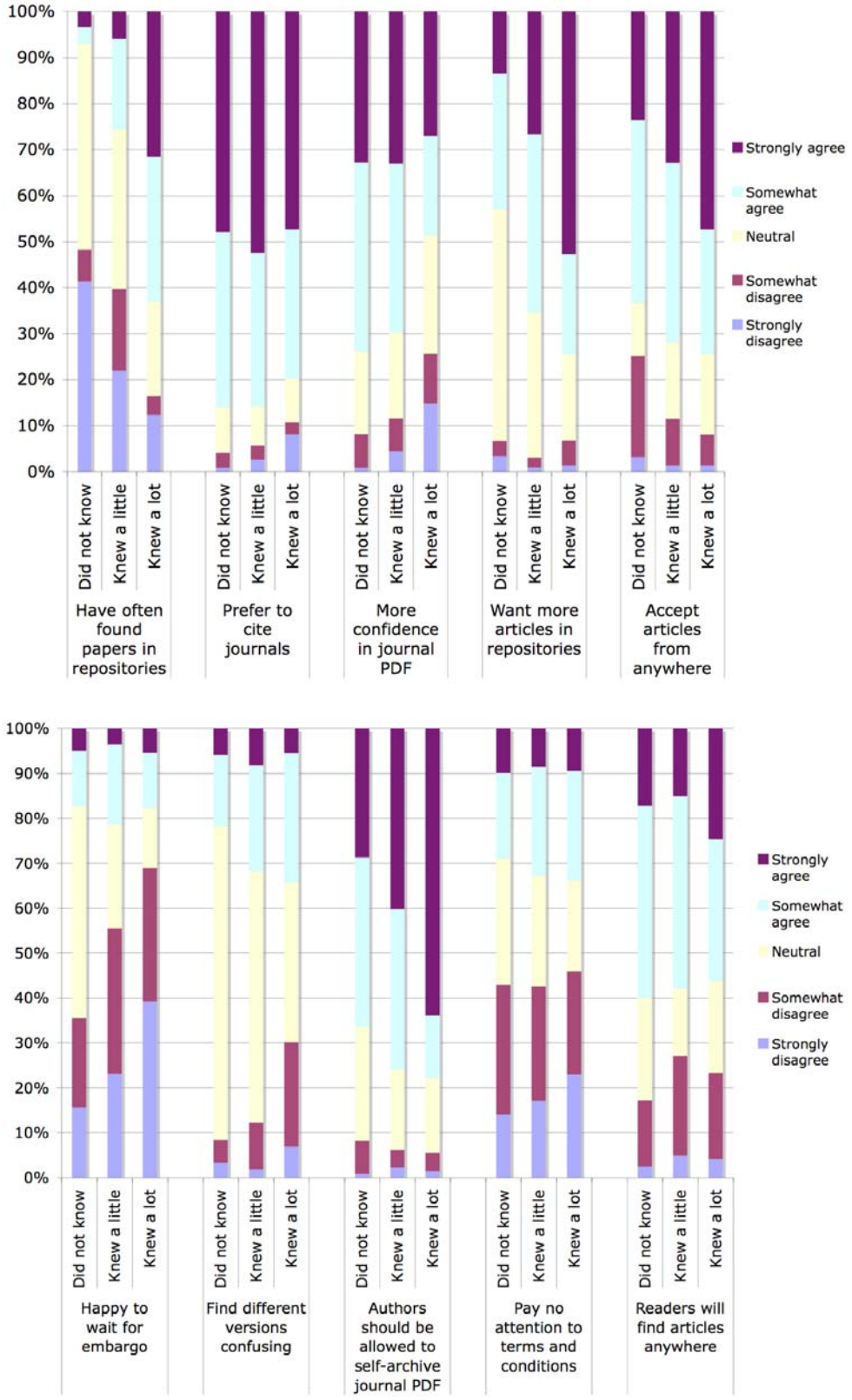


Figure 46: Effect of awareness of self-archiving on views of self-archiving

The response to the statement 'It is difficult to tell the different between pre-print (unrefereed) and peer-reviewed papers on repositories' was interesting. While most of those who did not know about self-

archiving, or who knew only a little, were neutral on this subject (presumably because they had not encountered different versions of the same paper), amongst those who 'knew a lot' both agreement and disagreement increased (χ^2=32.7, df=8, p<0.001). This possibly reflects different experiences, good or bad, with different versions of specific papers. While the small sample size means that it is not possible to investigate this further here, it could be that certain repositories are better designed than others in terms of differentiating between preprints and postprints.

Seniority of respondents, whether measured by number of years in the field or by role, had no significant effect on the responses to these questions. However, there was a tendency for respondents who had been in their field for a shorter period to be more likely to strongly agree that authors should be allowed to self-archive the journal PDF versions of their articles, as opposed to more established respondents who were more likely to only somewhat agree (χ^2=25.6, df=16, NS).

Respondents who have not used institutional repositories are more likely than those who have used institutional repositories to agree or strongly agree with the statement 'I would prefer to cite an article I'd found in a journal rather than one I'd found in a repository or on a website' (χ^2=156.2, df=4, p<0.001), and almost twice as likely to agree with the statement 'I have more confidence in the journal PDF than in the author's own version'. This indicates that the availability of the journal PDF is particularly important to these people and is possibly one factor in their choice of sources.

Author behaviour

A total of 349 respondents (81%) had published a paper, and two thirds of these (229) had done so in 2006 (Figure 47). Just under 70% of authors were responsible for submitting the paper to the journal. This was not affected by how recently the paper had been published (χ^2=1.7, df=5, NS), but the subject field had a considerable impact. Respondents in the arts and humanities were significantly more likely to be responsible for publication of their paper (χ^2=35.8, df=14, p>0.05) than those in the sciences. This is probably related to the tendency for science papers to have multiple authors, where arts and humanities papers tend to be written by a single author, or at most two or three.

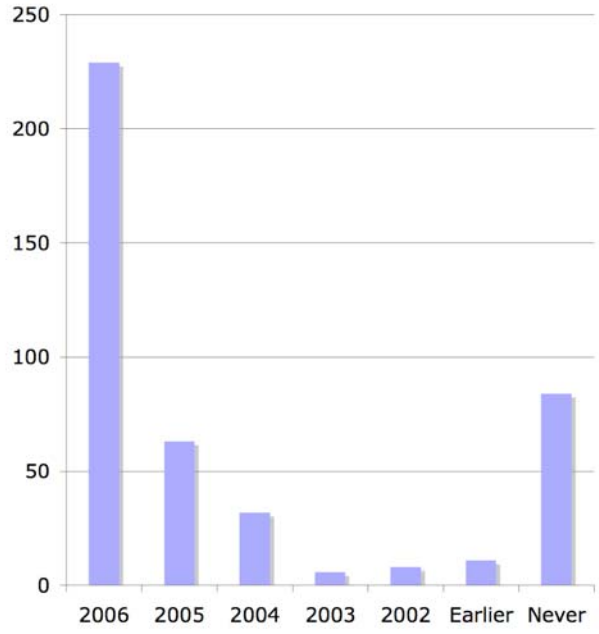


Figure 47: Publication year of last paper

Actual publisher policies

I asked respondents to name the journal that their most recently published paper appeared in. I then searched on the journal name in Google and on publisher websites to determine the publisher of each journal named (36 journals could not be traced in this way). From the publisher websites and from RoMEO I gathered as much information as I could about the policy of each publisher and journal with regard to copyright and self-archiving. There is some potential for inaccuracy here: it is possible that some respondents came to non-standard agreements with their publishers, and for those respondents whose last paper was published some time ago, the publishers' agreements may have been different at the time. However, for the most part, the information on publisher policies should be up to date, particularly as most of the respondents had published in the last year.

The publisher copyright policies are given in Figure 48. Most (62%) of the publishers used by respondents of this survey required a transfer of copyright from the author to the publisher, while many of the rest (16%) required an exclusive licence to publish; in effect, this also means that the publisher gains many of the privileges of copyright ownership, including the right to give or deny permission for publication elsewhere.

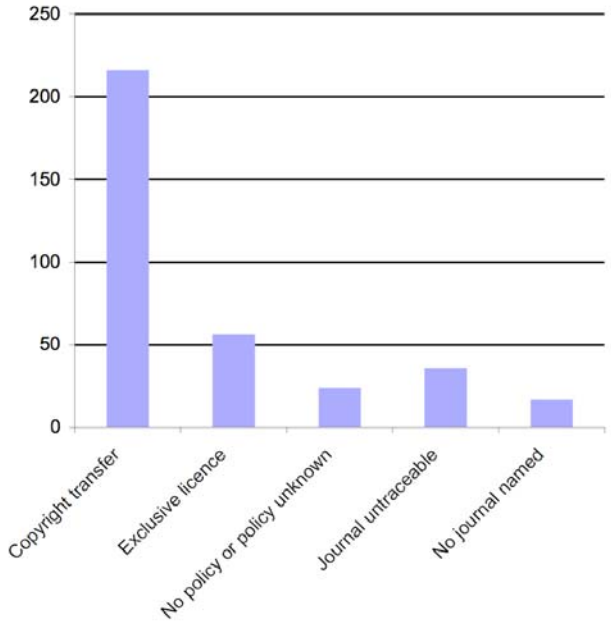


Figure 48: Copyright policies of publishers

Most publishers allowed some form of self-archiving of papers: 75% of respondents had published their most recent paper in a journal that allowed self-archiving to some degree. The majority allowed self-archiving of both preprints and postprints. A small number of respondents had published most recently in a journal that did not allow self-archiving at all, and a few respondents had published their most recent paper in an open access journal (Figure 49).

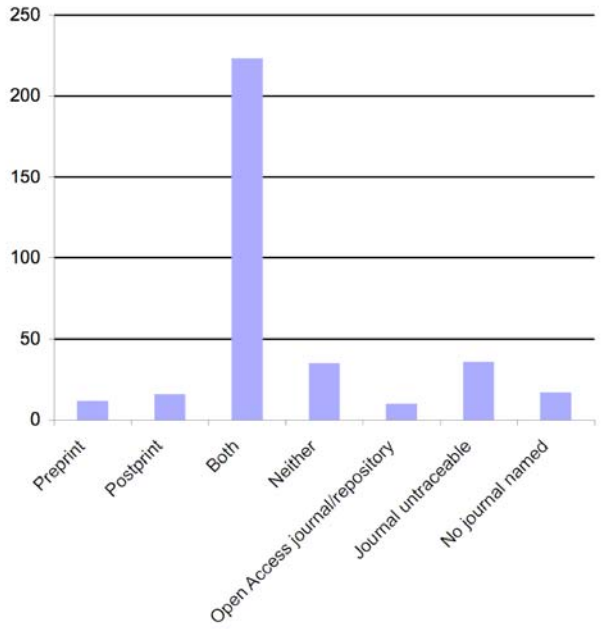


Figure 49: Self-archiving policies of publishers

Moreover, most respondents were publishing in journals that did not place any embargo on self-archiving (Figure 50). Of those that did have an embargo, 12 months was the most common length of time. This is, of course, the maximum time allowed by certain funders such as the Wellcome Trust and NIH.

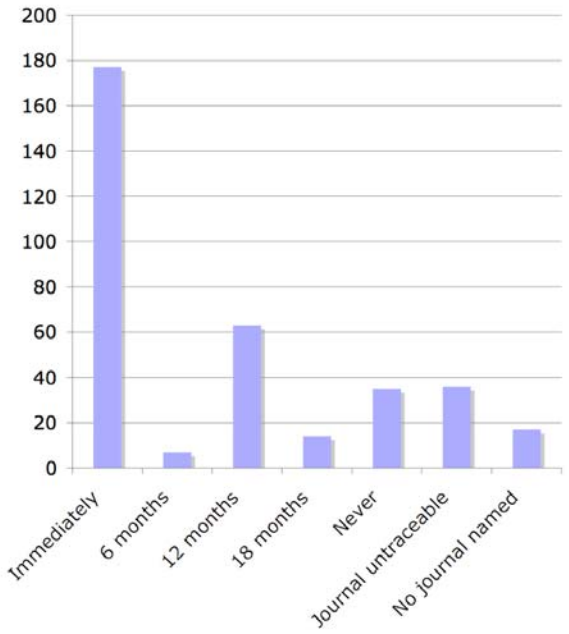


Figure 50: Time before an article can be self-archived

In addition, just over 30% of journals offered some form of open access publication, where articles were freely available from the publishers' websites (Figure 51). Thirteen authors (4%) had published their most recent article in a journal that was entirely open access, and nine (3% had published in a journal that made articles available after a certain period of time had elapsed, typically six months. The most common form of open access publication offered by journals in this survey was an author-pays option as an alternative to standard publication. Only one of the respondents in this survey explicitly stated that they had taken advantage of such an option (although admittedly they were not asked, this not being the focus of the survey) and so it can be assumed that the rest published under the standard model for their journal.

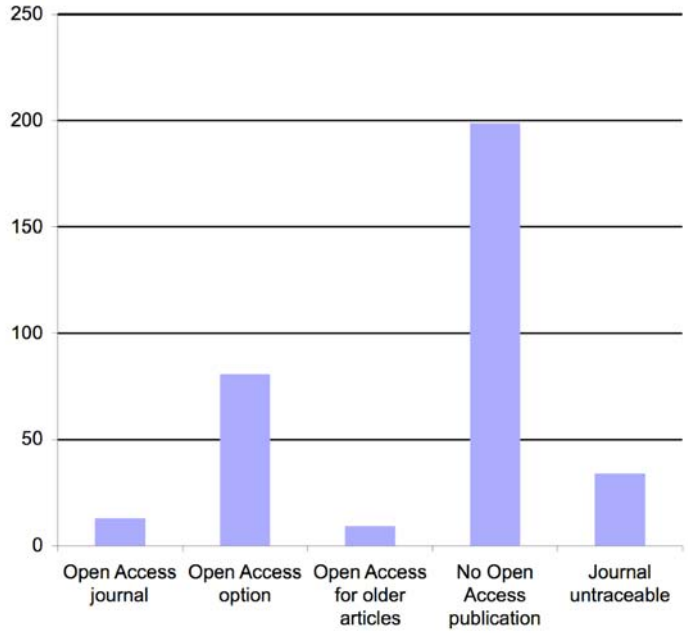


Figure 51: Open Access policies of publishers

Perception of publisher agreements

I asked authors what they thought they were allowed to do with their papers. Although I asked respondents to 'tick all that apply', most respondents only chose one option. Aggregating the first five responses gave an overall value of 94 respondents (27%) who believed that they were explicitly permitted to self-archive in some form (Figure 52). Over 50% of respondents who had published a paper, however, did not know what conditions were attached to its use. Respondents who were responsible for submitting the paper to the journal tended to be less likely to answer 'I don't know' to this question, although this was not significant (χ^2=3.8, df=1, NS). Respondents who answered 'I don't know' were also more likely to agree with the statement 'I don't pay much attention to publishers' terms and conditions when I publish a paper' (χ^2=16.1, df=8, p<0.01).

Those respondents who did not choose any of the options for permitted actions or the 'don't know' option can be taken to be those who actively believe that that are not allowed to self-archive their paper in any form. Forty-two respondents (12% of authors) fell into this category.

Respondents were less likely to answer 'Don't know' to this question if they had some awareness of self-archiving (χ^2=29.8, df=4, p<0.001). They were also more likely to believe that the journal they had published in allowed them to self-archive (Figure 53). This implies those who know about self-archiving

are actively researching the self-archiving policies of the journals they publish in.

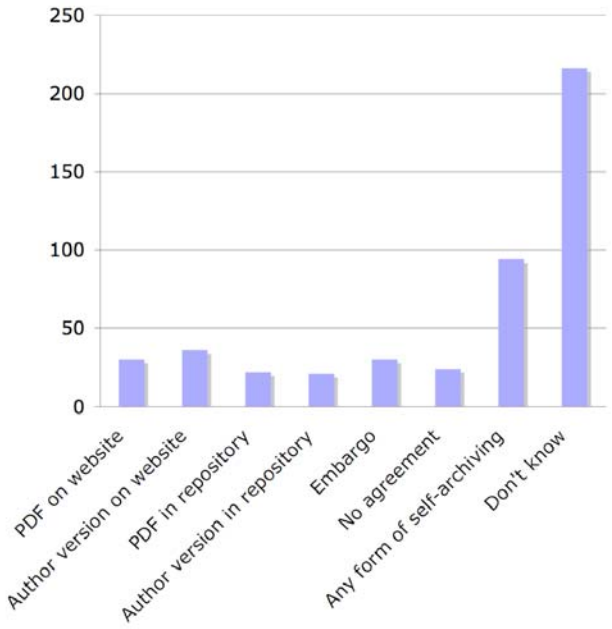


Figure 52: Author perception of permissions

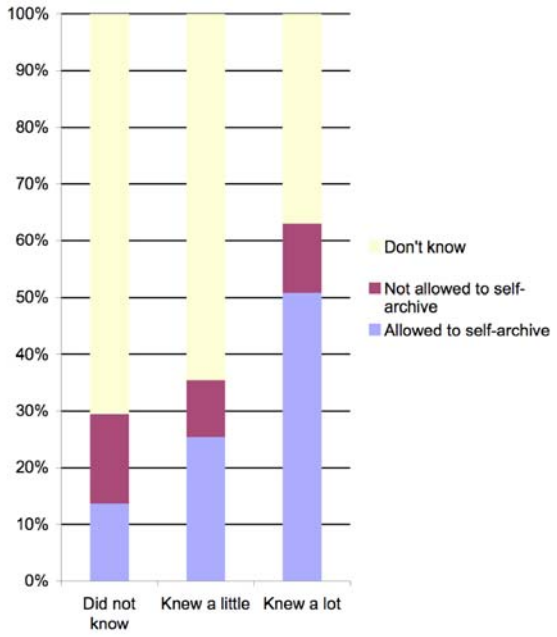


Figure 53: Authors' knowledge of self-archiving and perception of self-archiving permissions

However, when it came to actual journal policies, those who knew about self-archiving were no more likely to choose journals which allowed self-archiving than those who did not know (χ^2=1.3, df=4, NS). This is probably to some extent a reflection of the fact that most of the journals in this study had policies

allowing self-archiving, so that even those respondents who did not know about self-archiving and did not use journal self-archiving policy as a criterion in choosing a journal to publish in were still likely to publish in a journal which allowed some form of self-archiving.

There was some correlation between the actual policy of the publisher and the policy as perceived by the author, as shown in Figure 54. Of the respondents who did not choose 'don't know', those who published in journals which allowed self-archiving were twice as likely to believe that they were allowed to self-archive their paper than those who published in journals that did not (χ^2=7.9, df=2, p<0.05). However, 14% of respondents who published in a journal that did not allow self-archiving believed that it was allowed.

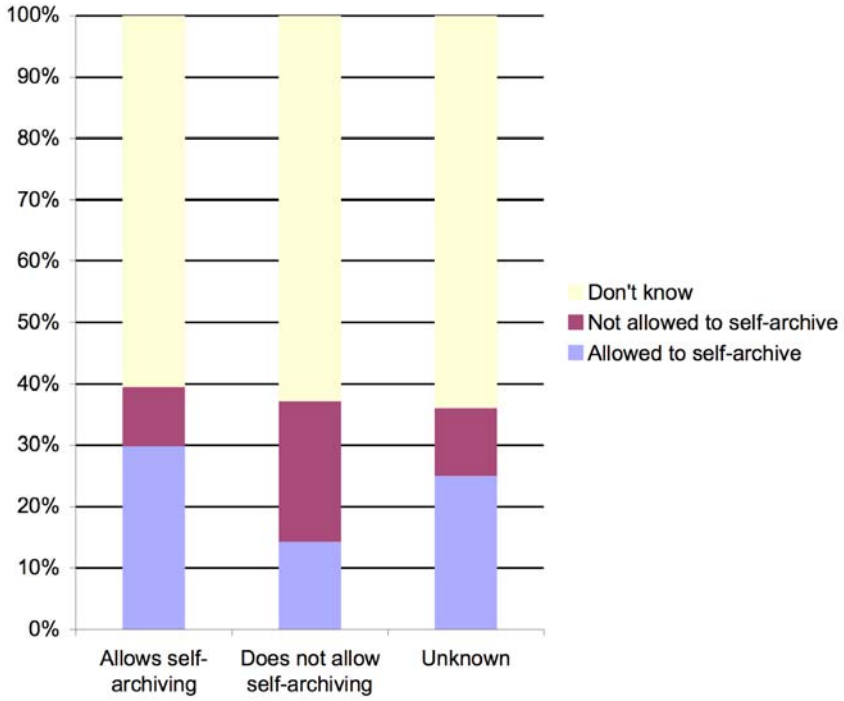


Figure 54: Accuracy of authors' perceptions of self-archiving permissions

Even those authors who knew a lot about self-archiving were unlikely to have accurate beliefs about the self-archiving policy of their journal (Figure 55), even though, as discussed earlier, they were less likely to say that they did not know. However, this is a very small sample: only seven respondents who 'knew a lot' about self-archiving had published their most recent paper in a journal which did not allow self-archiving, so it is probably not meaningful by itself that three of them (43%) believed that it did.

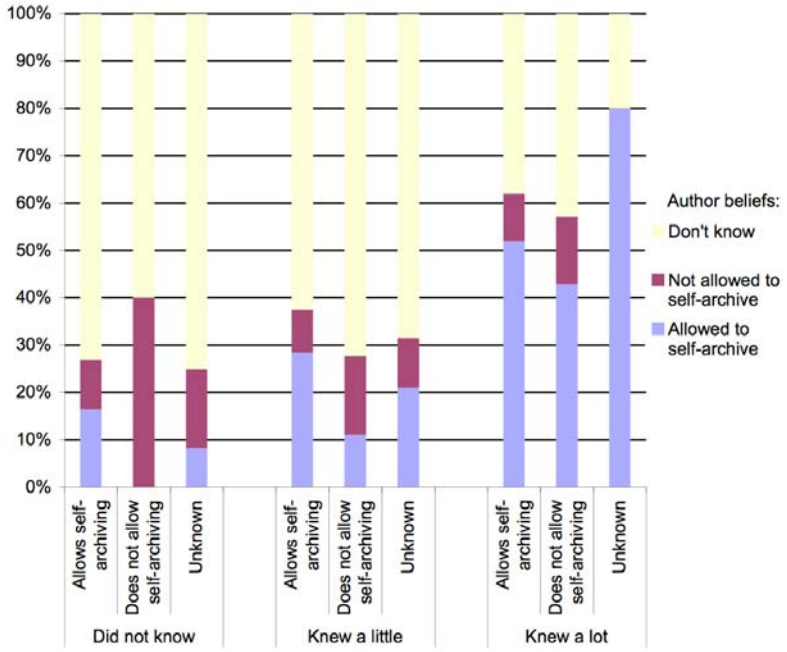


Figure 55: Effect of knowledge of self-archiving on accuracy of authors' perception of self-archiving policy

Most authors believed that their journal publisher owned the copyright for their article, even when this was untrue (Figure 56).

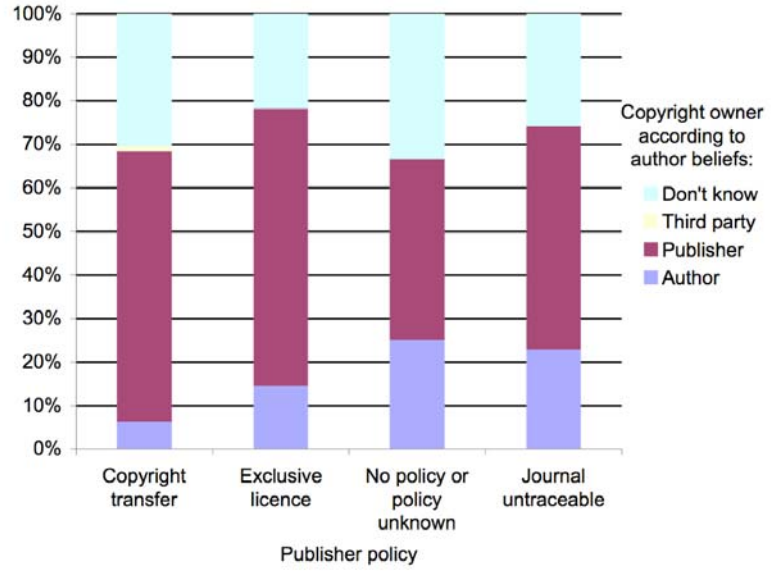


Figure 56: Accuracy of authors' perception of copyright ownership

Authors were also unlikely to accurately report their journal's embargo policy, and tended to assume there was no embargo regardless of the actual policy (Figure 57). There was no correlation between authors' beliefs and the actual policy of the journal publisher (χ^2=3.8, df=5, NS).

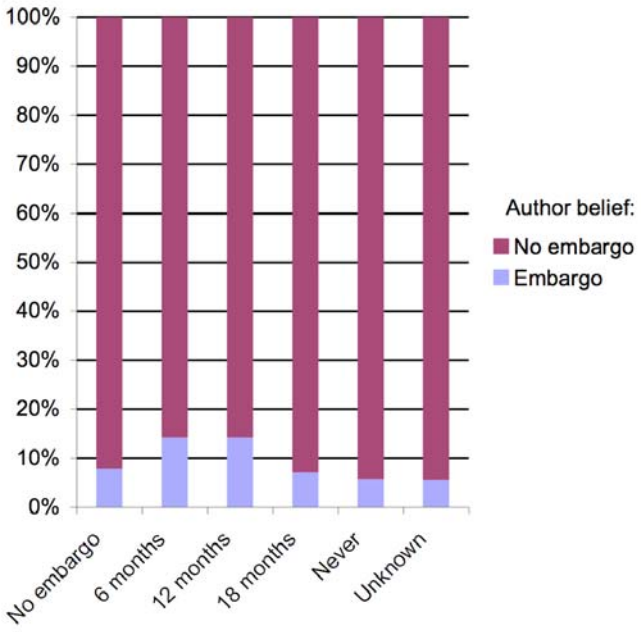


Figure 57: Accuracy of authors' perception of embargo policies

Self-archiving of most recently published paper

Overall, 96 respondents (28% of authors) said that they had self-archived or were intending to self-archive their most recently published paper in some form. Twenty-two (6%) said that they were waiting for an embargo to expire. The most popular form of self-archiving was uploading the journal PDF version to the author's website (Figure 58). The PDF was the more popular choice both for uploading to a personal website and for placing in a repository, and 52% of those who were placing their article in a repository were also placing it on a personal website. Fewer respondents chose options involving self-archiving their own versions, and of those who were placing their own version in a repository, 69% were also placing it on a personal website. There was little overlap between the two groups (7% were self-archiving both the PDF and the author version). In addition, twelve respondents mentioned that they would send offprints to other researchers on request:

> I always honor people's requests for PDFs that I receive by e-mail. I believe the vast majority of researchers do. This is, in other words, another form of repository that is not mentioned in this web questionnaire.

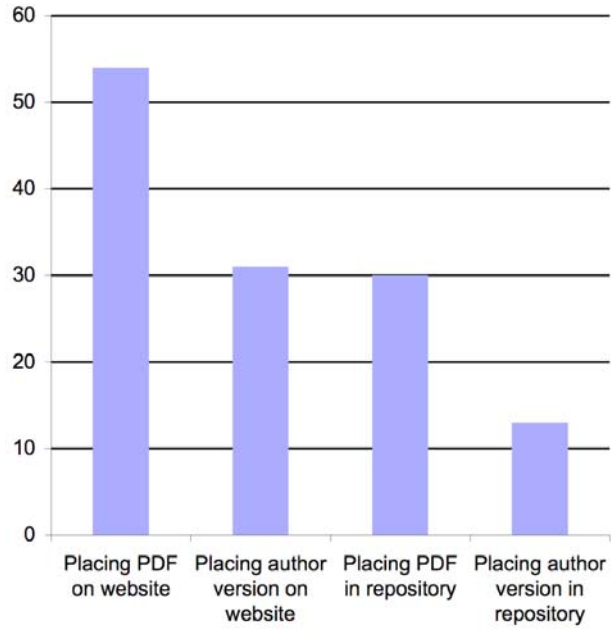


Figure 58: Methods of self-archiving of most recent paper

Figure 59 shows that, unsurprisingly, respondents who were more aware of self-archiving were more likely to self-archive their current paper than those who did not know about self-archiving (χ^2=51.6, df=2, p<0.001). Similarly, respondents in subjects where there was more awareness of self-archiving were more likely to self-archive their article than those in other subjects (χ^2=48.8, df=14, p<0.001).

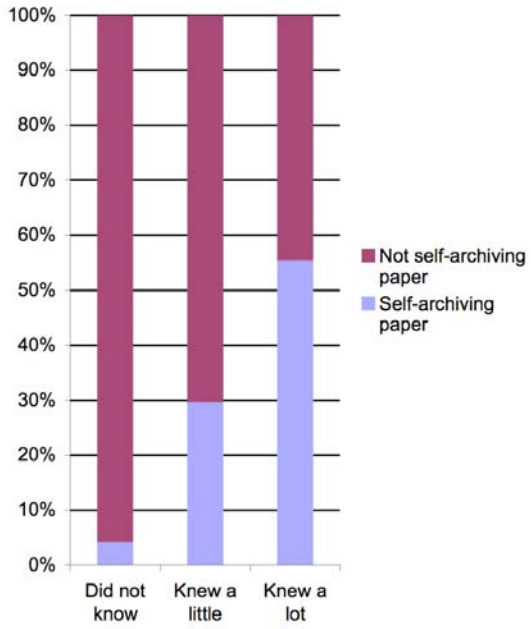


Figure 59: Effect of awareness of self-archiving on intention to self-archive current paper

Figure 60 shows the difference in attitudes between self-archivers and non-self-archivers. Respondents who were self-archiving their paper were less likely to agree, and more likely to disagree, with the statement 'I would prefer to cite an article I'd found in a journal rather than one I'd found in a repository or on a website' (χ^2=10.0, df=4, p<0.05). They were more likely to disagree with the statement 'I have more confidence in the journal PDF than in the author's own version' (χ^2=10.0, df=4, p<0.05). They were not, however, more likely to disagree with the statement 'I don't pay much attention to the publishers' terms and conditions when I publish a paper' (χ^2=0.7, df=4,NS).

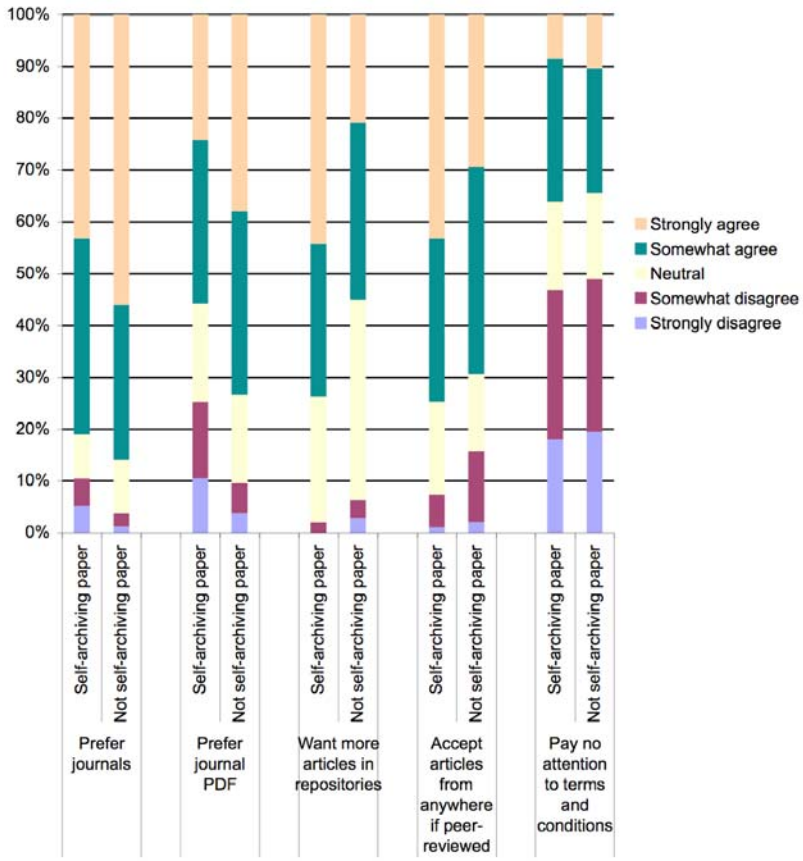


Figure 60: Attitudes and action of self-archiving

Reasons for and against self-archiving

The most popular reason for not self-archiving a paper was that respondents did not see the need for it (Figure 61). Two of the respondents who chose this option commented that this was because the journal was open access in any event, although, overall, the same proportion of respondents viewed self-archiving as unnecessary regardless of the open access policy of their journal (χ^2=1.0, df=4,NS). The second most common reason was lack of time:

> In principle, I approve of articles being publicly available on the Web. But I am just too busy to take the responsibility for doing this; it's just one extra burden in an already heavy workload.

I have better things to do with my time than faff around with copyright permissions and PDFs.

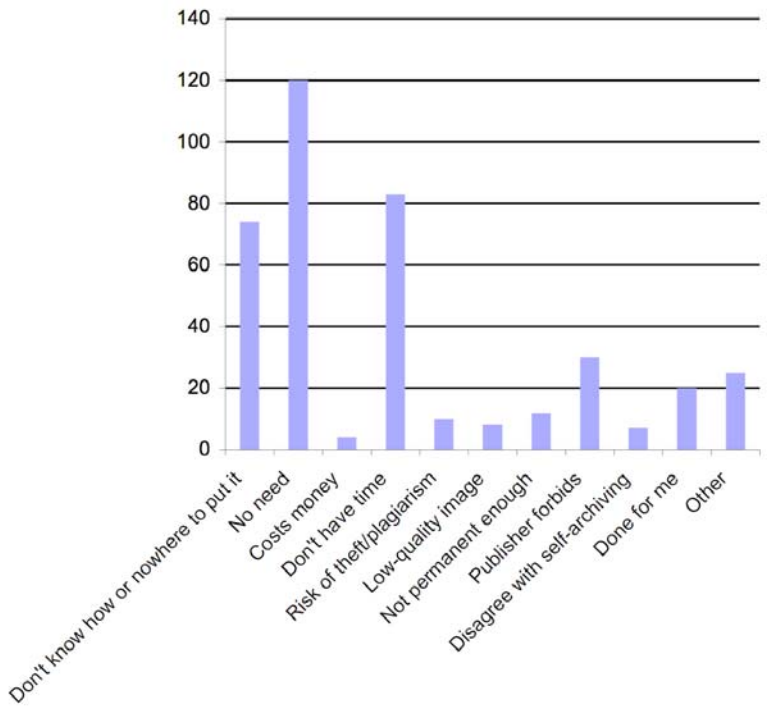


Figure 61: Reasons against self-archiving

The most common 'other' reason, stated by eight respondents, was that it had not occurred to them or they had not thought about it. Five were unsure of what they were allowed to do, and two were unsure of what had been done because they were not the lead author and so were not responsible for any self-archiving. Seven respondents disagreed with the idea in principle, with reasons ranging from 'All good articles are published in reputable journals to which University libraries have subscriptions' to the idea that searching for articles in repositories (or, indeed, elsewhere) meant that readers would not browse and so could not accidentally come across useful articles that they would not otherwise have read.

The most common reason for self-archiving, chosen by 113 respondents, was to 'make my work as accessible as possible'; in other words to increase access to the paper for its own sake (Figure 62). (This is not just a matter of increasing citations, which was offered as a separate option and chosen by 50 respondents, all but three of whom also chose increasing access as a reason.) Thirty-two of the respondents who chose this option had not stated in the previous question that they were self-archiving their current paper; it can be assumed that they were either referring to a paper that they had self-archived in the past or expressing a desire to self-archive in the future. One respondent commented 'It seems that it is becoming a standard, so I might as well do it'.

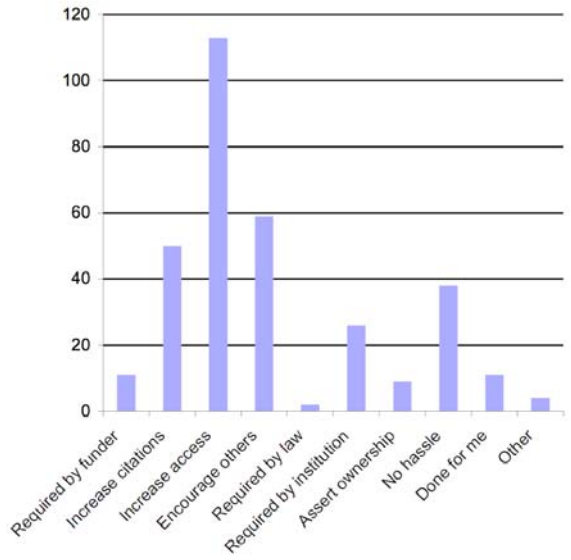


Figure 62: Reasons for self-archiving

Difference between self-archived and journal version

A total of 138 respondents had self-archived at least one of their papers in some form, a greater number than had self-archived or were planning to self-archive their most recent paper. Of those, 71 (57%) had made their most recently self-archived available in a form identical to the one available in the journal: in other words, they had self-archived the journal PDF (Figure 63). Another 30 (24% of self-archivers) had made available a version that was identical except in layout.

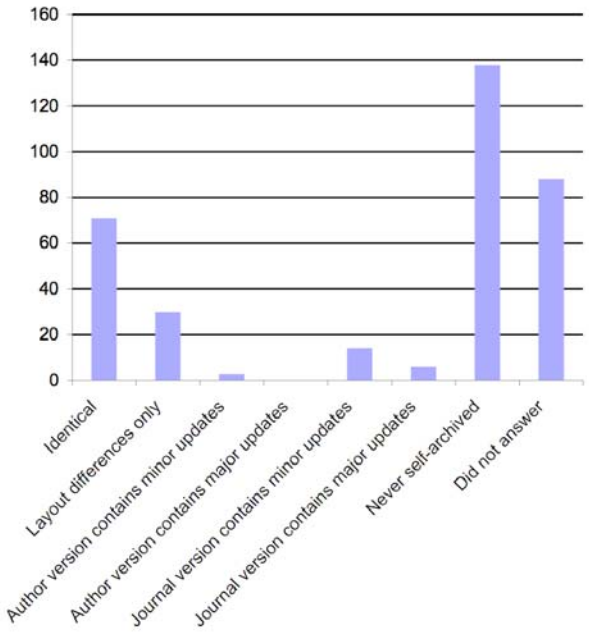


Figure 63: Difference between archived and journal version of last self-archived article

Six respondents reported that the journal version contained major corrections, i.e. the self-archived version was a preprint; the other authors who answered the question and had self-archived a paper all reported minor changes of the level of spelling and grammar corrections. No respondent reported making major changes to their own version that were not also in the journal version, although one respondent who chose the option of minor updates in the self-archived version commented that they included additional information that had had to be cut to comply with the journal's word limit. Most authors (81%) are therefore self-archiving a version of their article which is identical in content to the journal version, and most of the rest are self-archiving versions whose content is not greatly different from the point of view of scholarship. Although this is a small sample, it suggests that self-archived articles are indeed as trustworthy from the point of view of the reader as the journal versions.

Forty respondents who had self-archived their most recent paper as 'the final PDF' reported in this question that they had self-archived a version identical to the journal's. This accounted for 62% of all who were self-archiving their most recent paper. These authors, then, are definitely self-archiving the journal PDF version of their article. Of these, 26 (65%) had not reported that they had permission to use the journal PDF on their webpage or in a repository. In fact, according to the publishers' actual policies where these could be ascertained, 27 (68%) did not have permission to self-archive their journal PDF, even after any embargo had expired.

This indicates that these authors are self-archiving the journal PDF without regard to the publishers' terms and conditions, and some respondents were quite explicit about this:

> I had to sign a form of some sort giving away my rights to replicating my own work, but our institution still expects these publications to be put in our eprints repository (and it is). I don't know if this will get me in trouble, but I guess it would be more effort for the publisher to chase up so many individuals than it is worth.
>
> We have most likely uploaded the paper regardless of conditions to our repository.

Although this is too small a sample to claim a statistically significant relationship, and although the attitudes of authors are not the main subject of this survey, this is apparently the same phenomenon observed in Swan and Brown (2005), where authors who were unsure whether they were permitted to upload the journal PDF did so anyway, and Antelman (2006) where authors in the social sciences uploaded the final PDF regardless of conditions.

Format of uploaded papers

As with downloaded papers, PDF was by far the most common format for article upload (Figure 64), with almost all respondents who answered this question picking this option. For those respondents who are using repositories, this may be a reflection of the policies of their institutional repository, as many of them require or prefer PDFs. As noted above, in many cases these will be the journal PDF versions. In any event, this indicates that the papers are being self-archived in the readers' preferred format.

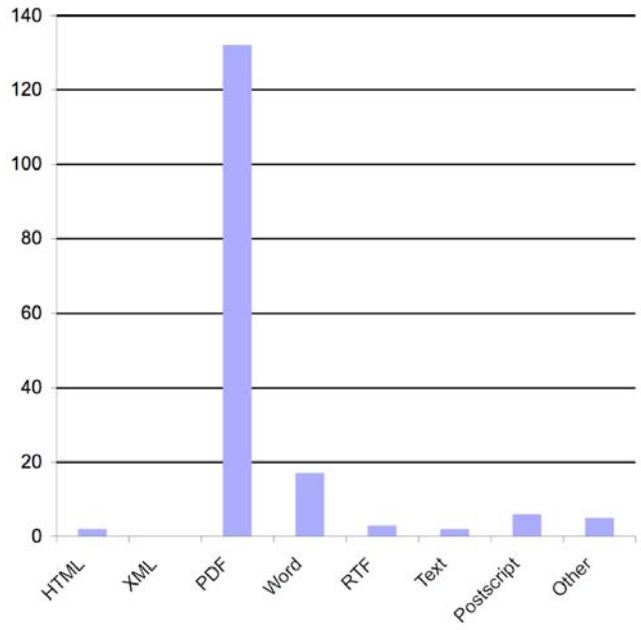


Figure 64: Preferred format for uploading self-archived papers

Comments

Most of the respondents who commented were positive about the idea of self-archiving, even if they had not heard of it before.

> This has been an interesting exercise. As a result I will definitely think about new means of dissemination and talk to colleagues about their practice.
>
> I would like to investigate more into the area of repository holding of articles. It is something that I have not come across.
>
> This has made me think of a topic about which I am vaguely aware, but need to know more! My impression is that in the Humanities we have not really considered these issues.
>
> My feeling is that where research is funded by the public, the results should be made freely available; in my view, this applies to much research carried out in UK universities.

Other respondents expressed themselves quite vehemently against the concept of self-archiving, for various reasons:

> The final printed (pdf) file is the ONLY protected and verified version that has the added value of peer-review and editorial assent. Post-prints are not acceptable and are a second-rate way of presenting material … If there were more OPEN ACCESS journals then the whole thing becomes totally irrelevant. I fully support OPEN ACCESS publication and consider that institutional or other repositories are a waste of time and effort.

> Making all this stuff easily accessible online is dispiriting. I don't WANT to know that there are 15,000 articles written about John Stuart Mill's concept of liberty. In the past, I could happily ignore 14,998 of them. Now, because they are available online, I cannot.

> It's a pointless process. All good articles are published in reputable journals to which University libraries have subscriptions.

Quality and ability to verify the material was a major concern, even where respondents were in favour of self-archiving in principle:

> Quality of science as reflected by stringent peer review is the most important factor in publishing scientific work. This is most obviously achieved by publishing in good journals. There is a danger with self archiving that quality will suffer.

> My main concern with papers found in repositories is peer reviewed status. I would not cite a paper that doesn't have peer reviewed status.

> It seems to me that the main issue is whether a published paper has been reviewed, and if so who is going to pay for that. I would want to have confidence that an article I read, even through Google or similar, has been properly reviewed and is likely to be bona fide.

> My concerns are two-fold: a) whatever I recommend, students should be able to easily access b) that information should be peer reviewed.

A few respondents, however, were simply confused by the whole issue:

> I still don't really understand how a repository works. As far as I know it is a standard term when publishing in a journal that the paper (in its exact form) may not be reproduced anywhere. I would be wary of upsetting editors when there are few relevant journals in my field!

Conclusions

Reader use of self-archived articles

Overall, 71% of respondents in this survey had used a self-archived paper, a greater proportion than formally knew about the concept of self-archiving. Most of these papers were from author websites, particularly among those who knew less about self-archiving; this reflects both the larger amount of content available in this way and the lack of knowledge of repositories as a source of material. It probably also reflects the widespread use of Google as a search tool, which, unlike Google Scholar, is more likely to return website-based articles than repository-based ones. Over half the respondents in this survey are already using search facilities that can find papers in repositories or on authors' personal websites; for 40% of respondents such a service is their preferred means of finding papers. However, they may well not be aware of the option for obtaining papers from repositories, or may not notice where their papers are coming from.

The prime factor in readers' use of self-archived articles was their knowledge of self-archiving, which is greater than shown in previous studies, suggesting that awareness is increasing over time. The majority of respondents had heard of self-archiving, and respondents who 'knew a lot' about self-archiving were over twice as likely to have used an article obtained from an author's website, and ten times more likely to have used an article from a repository, as respondents who 'did not know' about self-archiving, though some respondents may be using articles from repositories without knowing it.

However, this knowledge itself is affected by a number of factors. First, it depended on the subject field: respondents in the fields of mathematics and computer science were particularly likely to know about self-archiving, while those in medicine were particularly unlikely to. This agrees with the findings of Allen (2005) and Andrew (2003), who demonstrated similar trends in self-archiving behaviour of authors in these fields. It is possible that in medicine in particular the preference is for open access journals rather than self-archiving, and more research should be done to confirm this.

Second, it is likely to be related to advocacy efforts and advertising of self-archiving and repositories to researchers. Although awareness of self-archiving did not correlate strongly with the presence of an institutional repository at the respondent's institution, this may be because the repositories in several of the institutions were relatively new or because the effect of authorship or subject field is stronger. But the more articles a researcher had produced in the last year, the more likely they were to be aware of and have used self-archived articles, and respondents who had not published a paper were less likely to be aware of self-archiving. This suggests that publicising of repositories by institutions or funders such as RCUK, primarily directed towards authors in order to get them to upload their papers into the repository, means that end-users who do not publish papers are not being informed of self-archiving, and will continue to access articles via subscription. Admittedly, at this stage many repositories lack content: only half the institutional repositories involved in this study had more than a few hundred items; they are also not widely known to exist: over half the respondents to this survey did not know whether their own institution had one, even when it did. While respondents to this survey were more likely to

have self-archived their own papers if they were aware of the formal concept of self-archiving, it is unclear that the availability of repositories will naturally lead to the presence of content therein. Fried Foster and Gibbons (2005) suggest ways to encourage academic authors to deposit more of their articles in repositories.

Respondents who had spent less time in the field, or who had a more junior role, did appear to be slightly more likely than more senior researchers to have used self-archived material, though there was not a great difference, and there was no difference in awareness of self-archiving. Admittedly, this is a small sample: any effect of seniority might have been overshadowed by a much stronger subject field effect. The activity of respondents in terms of number of papers published seems to have a stronger effect than seniority in general.

Author self-archiving behaviour

Less than a third of authors had self-archived their most recent paper (or planned to), either in a repository or on their own website. The lack of self-archiving activity translates into a lack of self-archived material for people to access, which sets up a feedback loop which could explain some of the differences between the various subject fields. However, of self-archivers, 62% were self-archiving the journal PDF, with or without permission. This appears to be less due to a lack of knowledge or understanding of the publication terms and conditions than to do with these authors' feeling that the journal PDF is important (or easier for them to self-archive), and self-archiving is important (or mandated), and therefore they are deliberately self-archiving the journal PDF. It is probably impractical for journal publishers to monitor compliance to their conditions of publication and apply sanctions to all who do not comply; suing authors for breach of copyright would probably result only in very bad publicity for the publishers. Self-archiving of the journal PDF is therefore likely to continue, and possibly increase with the increased prevalence of self-archiving mandates by funders and institutions.

Perception of repository value

Despite the widespread use of self-archived articles, self-archiving is not yet widely accepted except in a few subject fields. The most common reason for using a self-archived paper is lack of access to the official version, suggesting that, currently at least, the subscribed version is valued more highly than the self-archived version. Readers prefer to use the journal citation in their papers, suggesting that the work of the journal publisher in terms of peer review is valued, and two thirds have more confidence in the journal PDF than in the author's own version: they would like more articles in repositories, but they want them to be journal PDFs.

However, respondents who were more aware of the phenomenon of self-archiving were less likely to place a high value on the journal PDF and more likely to use self-archived articles for convenience – in other words, in at least some cases they were choosing self-archived papers *in preference* to subscriber versions. They were more likely to consider repositories to be as good as journals, and less likely to consider personal websites to be untrustworthy sources of material.

Implications for scholarly publishing

The current situation appears for the moment to be one still favourable to journal publishers. The perception of the journal PDF as more authoritative means that, overall, researchers still prefer to obtain online articles from the official journal websites. However, with increased knowledge of self-archiving comes a decrease in perceived value of the official journal version over the author's own version of a paper, and an increased willingness to look elsewhere for articles, not just out of necessity but also just for convenience. At the same time, this same value assigned to the journal PDF, along with the convenience of having a pre-made document, means that authors are breaching their conditions of publication to self-archive the official journal PDF version of their article, meaning that much of the added value of accessing articles from publisher sources limited to subscribers only is lost.

Knowledge of self-archiving appears to have increased over the past few years and is likely to continue to increase over time, at least among academics. As more institutions develop repositories and promote them to their members; as more funders mandate open access publication of their articles and suggest self-archiving as a means of achieving this; as more papers become available to researchers through repositories, people will become aware of the concept of self-archiving in general and repositories in particular. Whether this will translate to a decrease in subscription is unclear: Ware's (2006) survey of librarians, who are usually responsible for deciding which journals to subscribe to, found that only 12% of librarians cited availability of the content of a journal in repositories as a factor in subscription cancellations, and almost 50% of librarians believed that the entire content of a journal would have to be available in this way before they would consider cancelling their subscription to that journal. Given the current state of institutional repositories, and the piecemeal manner in which they are populated (not by journal but by author), this will take a long time.

Ware (2004b) surveyed journal publishers on their opinion on the likely impact of self-archiving and found that while 44% agreed that it would have a substantial impact, 74% believed that this impact would be neutral, with advantages and disadvantages equally balanced. From the results of this survey, it would seem that in the short term little will change: too few articles are available in repositories, and too few people value them, for academic librarians to cancel subscriptions, while outside of academia too few people are aware of their existence. However, the change of attitudes that comes with increased knowledge of self-archiving means that in the long term, as repositories fill with content as mandated by institutions and funders, readers, who already do not generally care where the paper comes from as long they can have some guarantee of its authenticity, are likely to begin looking there for their articles. This will happen faster if authors continue to self-archive the journal PDF versions of their articles, as many are at present. Publishers therefore currently have a grace period in which to develop a response to the increase in self-archiving and use of self-archived articles.[11]

[11] Pinfield, 2004 suggests a number of different strategies for publishers to adapt to a world in which self-archiving is commonplace, proposing that publishers rethink their role and work with repositories to manage peer-review and provide copy-editing and typesetting services.

Limitations of the study

The small sample size and its problems have been discussed already. In addition, the population was limited in that it consisted almost entirely of academics, a large proportion of whom were published authors; moreover, the limited number of institutions resulting from the way the sample was chosen may also have had an effect. However, the agreement between this survey and previously published reports indicates that these effects are small.

The survey did not ask respondents how often they had used self-archived articles or self-archived their own articles, and thus did not distinguish between people who did this regularly and people who had only done it once or twice. This distinction might have given an insight into how habitual the use of self-archived material is for researchers.

Some respondents also commented that it was difficult for them to answer some of the questions because little mention was made of 'grey literature': technical reports, presentations, and the like, which are not generally published in journals and are not necessarily peer-reviewed. I did not include these in the study as I did not feel that it was relevant to the question of how journal publishers would be affected, but it may be that this omission has biased the results somewhat in that it is not possible to tell whether respondents who have accessed self-archived material were accessing articles that had been published in journals, or whether they were accessing material whose availability would not affect journal publishing in any event.

Further research

This study suggests that end-users are not being made aware of self-archiving and repositories as a means for accessing research articles. However, the population of pure end-users was quite small in this survey, as the sample consisted mainly of academics who were mostly authors as well as readers. A similar survey should be carried out using a more mixed population including professionals and commercial researchers, who are suspected to use more papers and write fewer than academics.

This study also highlighted the large variation in self-archiving awareness and attitudes between different fields. While this may be a historical accident – some fields happen to have developed subject-based repositories and thus have a history of self-archiving, while others do not – there are some suggestions that there are cultural issues here, with the difference in valuing of self-archiving being related to differences in valuing of peer-review and speed of publication. In particular, it seems that in medicine the author-pays method of open access is favoured over self-archiving. More research is required into the difference between subject fields to determine how great these cultural differences are and how they affect attitudes to self-archiving and use of self-archived articles.

In order to monitor the awareness of self-archiving and to determine whether it does indeed continue to increase over the next few years, as seems likely, a similar survey to this one should be carried out in a few years' time, when more institutions would be expected to have repositories and be publicising them to their researchers.

References

Albanese, A. (2005). Blackwell offers open access plan. *Library Journal*, *130(6)*, 20.

Albanese, A. (2004a). Publishers protest at NIH. *Library Journal*, *129(14)*, 17.

Albanese, A. (2004b). Report: open access cheaper. *Library Journal*, *129(10)*, 24.

Albanese, A. R. (2005). Life after the NIH. *Library Journal*, *130(7)*, 56.

Allen, J. (2005) *Interdisciplinary differences in attitudes towards deposit in institutional repositories*. Masters dissertation. Manchester Metropolitan University, Manchester. http://eprints.rclis.org/archive/00005180/

Anderson, R. (2004). Author disincentives and open access. *Serials Review*, *30(4)*, 288-291.

Andrew, T. (2003). Trends in self-posting of research material online by academic staff. *Ariadne, 37,* http://www.ariadne.ac.uk/issue37/andrew/ (accessed 29 September 2006).

Anon. (2004). Open access: Publicly funded research results should be available to all. Financial Times Nov 9, 2004 p.18.

Anon. (2005). Science and technology: who pays the piper…; scientific publishing. *The Economist*, *374(8413)*, 82.

Antelman, K. (2006). Self-archiving practice and the influence of publisher policies in the social sciences. *Learned Publishing*, *19*, 85–95.

Ashburner, M., Ashmore, J., & al., e. (2005). Open letter from Fellows of the Royal Society regarding the Society's position statement on open access. http://www.frsopenletter.org/ (accessed 29 September 2006).

Awre, C. (2003). Open access and the impact on publishing and purchasing. *Serials*, *16*, 205–208.

Baldwin, C. (2004). What do societies do with their publishing surpluses? ALPSP and Blackwell Survey. http://www.alpsp.org/publications/pub8.htm (accessed 20 January 2006).

Banks, P. (2004). Open access: a medical association perspective. *Learned Publishing*, *17*, 135-142.

Belew, R. K. (2005). Scientific impact quantity and quality: analysis of two sources of bibliographics data. http://www.cogsci.ucsd.edu/~rik/papers/belew05-iqq.pdf (accessed 3 October 2006).

Bernard Wysocki, Jr (2004). Journals Resist Free Access To Medical Data. Wall Street Journal Oct 28, 2004 p.B.1.

Bjork, B.-C. (2005). A lifecycle model of the scientific communication process. *Learned Publishing*, *18*, 165-176.

Bosc, H., & Harnad, S. (2005). In a paperless world a new role for academic libraries: providing open access. *Learned Publishing*, *18*, 95-100.

Bosch, S. (2005). Buy, build, or lease: managing serials for scholarly communications. *Serials Review*, *31(2)*, 107-115.

Byrd, G. D., Bader, S.A., & Mazzaschi, A.J. (2005). The status of open access publishing by academic societies. *Journal of the Medical Library Association*, *93*, 423–424.

Carr, L., & Harnad, S. (2005). Keystroke Economy: A Study of the Time and Effort Involved in Self-Archiving. www.ecs.soton.ac.uk/~harnad/Temp/Keystroke3edited.doc (accessed 12 January 2006).

Chesler, A. (2004). Open access: a review of an emerging phenomenon. *Serials Review*, *30(4)*, 292-297.

Clark, N., & Scherlen, A. (2005). The balance point: looking at journals from both sides. *Serials Review,* doi: 10.1016/j.serrev.2005.08.003

Cockerill, M. (2004). The benefits to scientists of open access scientific publishing provided by the internet are too significant to be ignored,. Financial Times Nov 26, 2004 p.13.

Conference on Open Access to Knowledge in the Sciences and Humanities. (2003). Berlin declaration on open access to knowledge in the sciences and humanities. http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf

Cookson, C. (2005). Scientists reignite open access debate: publicly funded research. Financial Times, Aug 31, 2005, p.4.

Costello, A., & Osrin, D. (2004). Open-access publishing. *The Lancet*, *364(9428)*, 25-26.

Crawford, B. (2003). Open-access publishing: where is the value? *The Lancet*, *362(9395)*, 1578–1580. doi: 10.1016/S0140-6736(03)14749-6

Crawford, W. (2005). Open access and survivable libraries. *EContent*, *28(6)*, 42.

Daoutis, C., & Price, K. (2006) *The development of an institutional repository at the University of Surrey: a feasability study.* Report. University of Surrey, Surrey.

Davis, P., Ehling, T., Habicht, O., How, S., Saylor, J.M., & Walker, K. (2004). *Report of the CUL Task Force on Open Access Publishing Presented to the Cornell University Library Management Team August 9, 2004.* Cornell University Library. Technical Reports and Papers.

Butler, D. (2003). Scientific publishing: who will pay for open access? *Nature*, *425*, 554–555. doi: 10.1038/425554a

Doyle, B. (2005). Opening access to content. *EContent*, *28(6)*, 28.

Ellen, P. (2004). Warning: open access may be hazardous to the health of your scientific society. *BioScience*, *54*, 1060-1063.

Enderby, J. (2004). Internet upstarts v traditional publishers. Financial Times Nov 26, 2004 p.13.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology*, *4*, 692–698. doi: 10.1371/journal.pbio.0040157

Frank, M., Reich, M., & Ra'anan, A. (2004). A not-for-profit publisher's perspective on open access. *Serials Review*, *30(4)*, 281-287.

Fried Foster, N., & Gibbons, S. (2005). Understanding faculty to improve content recruitment for institutional repositories. *D-Lib Magazine, 11,* http://dlib.org/dlib/january05/foster/01foster.html (accessed 29 September 2006).

Friend, F. J. (2004). How can there be open access to journal articles? *Serials*, *17*, 37–40.

Gadd, E., Oppenheim, C., & Probets, S. (2003a). RoMEO studies 2: how academics want to protect their open-access research papers. *Journal of Information Science*, *29(5)*, 333–356. doi: 10.1177/01655515030295002

Gadd, E., Oppenheim, C., & Probets, S. (2003b). RoMEO studies 3: how academics expect to use open access research papers. *Journal of Librarianship and Information Science*, *35(3)*, 171-187.

Gadd, E., Oppenheim, C., & Probets, S. (2003c). RoMEO studies 4: an analysis of journal publishers' copyright agreements. *Learned Publishing*, *16*, 293-308. doi: 10.1087/095315103322422053

Gannon, F. (2005). Open access: scientists as paradoxical consumers. *Learned Publishing*, *18*, 295-299.

Gass, A. (2005). Paying to free science: costs of publication as costs of research. *Serials Review*, *31(2)*, 103-106.

Gedye, R. (2004). Open access is only part of the story. *Serials Review*, *30(4)*, 271-274.

Goodman, D. (2004). The criteria for open access. *Serials Review*, *30(4)*, 258-270.

Goodman, D. (2005). Open access: what comes next? *Learned Publishing*, *18*, 13-23.

Goodman, D., & Foster, C. (2004). Special focus on open access: issues, ideas, and impact. *Serials Review*, *30(4)*, 257-257.

Gregory, P. (2004). Open-access model to fund scholarly publishing has failed in the past. Financial Times Jun 7, 2004 p.16.

Guédon, J.-C. (2004). The "green" and "gold" roads to open access: the case for mixing and matching. *Serials Review*, *30(4)*, 315-328.

Hajjem, C., Harnad, S., & Gingras, Y. (2005). Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. *IEEE Data Engineering Bulletin.* Cited in http://eprints.ecs.soton.ac.uk/11688/ (accessed 2 October 2006)

Ham, M. F., Namtini, S. S., Luo, W.-J., & Xu, J.-X. (2004). Open-access publishing. *The Lancet*, *364(9428)*, 24-25.

Hangarter, R. (2005). Worrisome Profit Grab in Scholarly Publishing. Wall Street Journal Jun 21, 2005 p.A.17.

Harnad, S. (2001). Research access, impact and assessment. *Times Higher Education Supplement*, *1487*, 16. http://www.ecs.soton.ac.uk/~harnad/Tp/thes1.html

Harnad, S. (2003). Open access to peer-reviewed research through author/institution self-archiving: maximizing research impact by maximizing online access. *Journal of Postgraduate Medicine*, *49*, 337–342.

Harnad, S. (2005a). Fast-Forward on the Green Road to Open Access: The Case Against Mixing Up Green and Gold. *Ariadne, 42.* http://www.ariadne.ac.uk/issue42/harnad/ (accessed 23 November 2005).

Harnad, S. (2005b). Southampton workshop on UK institutional open access repositories. *Serials Review*, *31*, 175–177.

Harnad, S. (2005). Southampton workshop tackles open access. *Information Today*, *22(3)*, 1.

Harnad, S., & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine, 10,* http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/june04/harnad/06harnad.html (accessed 26 November 2005).

Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Stamerjohanns, H., & Hilf, E., R. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, *30(4)*, 310-314.

Harris, E. (2005). Institutional repositories: is the open access door half open or half shut? *Learned Publishing*, *18*, 85-90.

Hawkins, D. T. (2005). Does open access publishing really work in practice? *Information Today*, *22(1)*, 40.

Hernon, P., & Schwartz, C. (2005). Open access to research. *Library & Information Science Research*, *27(2)*, 139-140.

Hey, J. M. N., Simpson, P., & Carr, L. A. (2005). *The TARDis route map to Open Access: developing an institutional repository model*. Paper presented at the ELPUB 2005 Conference on Electronic Publishing, Kath. Univ. Leuven, 2005.

House of Commons Science and Technology Committee. (2004). Scientific publications: free for all? The Stationery Office, London.

Hubbard, C., Tamber, P., S, Lin, V., Furler, L., & Duckett, S. (2005). The implications of open access publishing for the medical community. *Evidence-Based Healthcare and Public health*, *9(6)*, 371-373.

Stevenson, I. (2004). Is open access the new vanity publishing? *Learned Publishing*, *17*, 83-84.

IRRA Project. (2006). IRRA: Institutional Repositories for Research Assessment: A JISC Project. http://irra.eprints.org/about.html (accessed 29 September 2006).

Jacsó, P. (2003). Proxy Searching of Non-Searchable and Poorly Searchable Open Access Archives of Digital Scholarly Journals. In *Digital Libraries: Technology and Management of Indigenous Knowledge for Global Access*. (pp. 552–555). Berlin, Germany: Springer.

Jacsó, P. (2005). Google Scholar and *The Scientist*. http://www2.hawaii.edu/~jacso/extra/gs/ (accessed 3 October 2006).

JISC. (2005a). Open Access. http://www.jisc.ac.uk/uploaded_documents/JISC-BP-OpenAccess-v1-final.pdf (accessed 30 November 2005).

JISC. (2005b). Digital repositories: helping universities and colleges. http://www.jisc.ac.uk/uploaded_documents/JISC-BP-Repository(HE)-v1-final.pdf (accessed 30 November 2005).

Willinsky, J. (2003). Scholarly associations and the economic viability of open access. *Journal of Digital Information, 4(2),* http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Willinsky/ (accessed 26 November 2005)

Jones, M. (2003). Archiving e-journals consultancy – final report. http://www.jisc.ac.uk/uploaded_documents/ejournalsfinal.pdf (accessed 2 October 2006).

Jongejan, A. (2004). The formula works, so don't tinker with it. Financial Times May 26, 2004 p.16.

Hunter, K. (2005). Critical issues in the development of STM journal publishing. *Learned Publishing*, *18*, 51-55.

Kaufmann-Wills Group. (2005). The facts about Open Access. http://www.alpsp.org/publications/pub11.htm (accessed 20 January 2006).

Kim, J. (2005). *Finding Documents in a Digital Institutional Repository: DSpace and Eprints*. Paper presented at the 42 Proceedings 68th Annual Meeting of the American Society for Information Science and Technology (ASIST), Charlotte, NC, USA, 2005.

King, D. (2004). *Who is going to pay for Open Access publishing?* Paper presented at the Cologne Summit on Open Access Publishing, Cologne, 2004. http://www.zbmed.de/summit/PPking.pdf (accessed 26 November 2006).

Kirsop, B., & Chan, L. (2005). Transforming access to research literature for developing countries. *Serials Review,* doi:10.1016/j.serrev.2005.09.003

Kousha, K., & Thelwall, M. (2006). *Google Scholar Citations and Google Web/URL Citations: A Multi-Discipline Exploratory Analysis*. Paper presented at the Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting, Nancy, France, 2006. http://eprints.rclis.org/archive/00006416/

Kurtz, M.J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., & Murray, S.S. (2005). The effect of use and access on citations. *Information Processing & Management*, *41(6)*, 1395-1402.

Lafferty, M. (2003). Making scholarly communication accessible: the Budapest Open Access Initiative. http://www.unc.edu/~melaffer/BOAI/BOAI.pdf (accessed 4 October 2006)

Lamb, C. (2004). Open access publishing models: opportunity or threat to scholarly and academic publishers? *Learned Publishing*, *17*, 143-150.

Law, S.M., Lowry, R. E., Motherswell, M.G., & Cody, S.M. (no date) The costs of electronic publishing of scholarly works. http://www.ryerson.ca/gcm/cep/THE_COSTS_OF_ELECTRONIC_PUBLISHING.pdf (accessed 2 October 2006)

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, *411*, 521–522. doi: 10.1038/35079151

Liesegang, T.J., Schachat, A., P., & Albert, D.M. (2005). The open access initiative in scientific and biomedical publishing: fourth in the series on editorship. *American Journal of Ophthalmology*, *139(1)*, 156-167.

Langston, L. (1996). Scholarly Communication and Electronic Publication: Implications for Research, Advancement, and Promotion. http://www.library.ucsb.edu/untangle/langston.html (accessed 29 November 2005).

MacCallum, C. J., & Parthasarathy, H. (2006). Open access increases citation rate. *PLoS Biology*, *5*, 661–662. doi: 10.1371/journal.pbio.0040176

Macleod, D. (2005). Royal society attacked for 'negative' open access stance. *Guardian Unlimited,* http://education.guardian.co.uk/higher/research/story/0,9865,1661107,00.html (accessed 29 September 2006).

McCabe, M., & Snyder, C. (2004). The economics of open-access journals. www.prism.gatech.edu/~mm284/OA.pdf (accessed 2 October 2006).

McVeigh, M. E. (2004). Open Access Journals in the ISI Citation Databases: Analysis of Impact Factors and Citation Patterns. http://www.isinet.com/media/presentrep/essayspdf/openaccesscitations2.pdf (accessed 30 November 2005).

Milmo, D. (2005). Lancet publisher scathing about university funding in UK. The Guardian Feb 18, 2005 p.23.

Montonen, C. (2005). The european physics publications scene: avant-garde and traditionalism. *Learned Publishing*, *18*, 149–151.

Morris, S. (2003). Open publishing. *Learned Publishing*, *16*, 171-176.

Morris, S. (2004a). *Open Access publishing and not-for-profit publishers*. Paper presented at the Cologne Summit on Open Access Publishing, Cologne, 2004.

Morris, S. (2004b). Open access: how are publishers reacting? *Serials Review*, *30(4)*, 304-307.

Morris, S. (2005). The true costs of scholarly journal publishing. *Learned Publishing*, *18*, 115-126.

Morris, S. (2006). Personal view: when is a journal not a journal? a closer look at the doaj. *Learned Publishing*, *19*, 1–3.

Mort, D. (2004). STM information market still looks healthy. http://www.researchinformation.info/special2004overview.html (accessed 2 October 2006).

National Institutes of Health. (2005). Final NIH Public Access Policy Implementation. http://publicaccess.nih.gov/publicaccess_imp.htm (accessed 29 November 2005).

National Library of Medicine. (2004). Fact Sheet: What's the Difference Between MEDLINE® and PubMed®? http://www.nlm.nih.gov/pubs/factsheets/dif_med_pub.html (accessed 29 September 2006).

Nicholas, D., Huntington, P., & Rowlands, I. (2005). Open access journal publishing: the views of some of the world's senior authors. *Journal of Documentation*, *61*, 497–519. doi: 10.1108/00220410510607499

Nicholas, D., Huntington, P., & Jamali, H. R. (2006). The impact of open access publishing on use and users: A deep log analysis of 'Nucleic Acids Research'. http://www.oxfordjournals.org/news/oa_report.pdf (accessed 29 September 2006).

Nicholas, D., & Rowlands, I. (2005). Open access publishing: the evidence from the authors. *The Journal of Academic Librarianship*, *31(3)*, 179-181.

OCLC. (2003). Five-year information format trends. http://www.oclcpica.org/?id=1235&ln=uk (accessed 2 October 2006).

Office of Fair Trading. (2002). The market for scientific, technical and medical journals. http://www.oft.gov.uk/NR/rdonlyres/A56C7602-C0BD-428D-BED2-36784363243B/0/oft396.pdf (accessed 2 October 2006).

Ojala, M. (2005). Open access: open sesame or opening Pandora's box? *EContent*, *28(6)*, 30.

Open Society Institute (2001). Budapest Open Access Initiative. http://www.soros.org/openaccess/read.shtml (accessed 19 January 2006)

Patterson, V., Barbour, C., & Mccallum, M. (2004). Opening access to scholarly papers. Financial Times Jun 2, 2004 p.18.

Peek, R. (2005). Counting OA journals. *Information Today*, *22*, 17–18.

Peek, R. (2004a). The free-access debate flourishes. *Information Today*, *21(5)*, 17.

Peek, R. (2004b). The cascading effect. *Information Today*, *21(8)*, 17.

Pentz, E. (2005). Crossref search pilot. *CrossRef Newsletter.* Cited in http://www.earlham.edu/~peters/fos/2005_02_13_fosblogarchive.html (accessed 3 October 2006).

Suber, P. (2003). Bethesda statement on open access publishing. http://www.earlham.edu/~peters/fos/bethesda.htm (accessed 27 November 2005).

Pinfield, S. (2004). Self-archiving publications. In Gorman, G. E., & Rowland, F. (Eds.), *International Yearbook of Library and Information Management 2004/2005: Scholarly Publishing in an Electronic Era*. (pp. 118–145) Facet Publishing. http://eprints.nottingham.ac.uk/archive/00000142/

Pinfield, S. (2005). A mandate to self archive? The role of open access institutional repositories. *Serials: the Journal for the Serials Community*, *18*, 30–34.

Poynder, R. (2004). No gain without pain. *Information Today*, *21(10)*, 1–5.

Poynder, R. (2004). Ten years after. *Information Today*, *21 (9)*, 1–5.

Prosser, D. C. (2003). From here to there: a proposed mechanism for transforming journals from closed to open access. *Learned Publishing*, *16*, 163-166.

Public Library of Science. (2004). Publishing open-access journals. http://www.plos.org/downloads/oa_whitepaper.pdf (accessed 2 October 2006).

Quint, B. (2004). Future of the NIH open access policy. *Information Today*, *21(9)*, 7.

RCUK. (2006). RCUK position statement on access to research outputs. Research Councils UK. http://www.rcuk.ac.uk/access/2006statement.pdf (accessed 4 October 2006)

Regazzi, J. (2004). The shifting sands of open access publishing, a publisher's view. *Serials Review*, *30(4)*, 275-280.

Regazzi, J., & Caliguiri, N. A. (2006). Publisher and author partnerships: a changing landscape. *Learned Publishing*, *19*, 183–192.

Reid, C. (2004). STM publishers fight open access proposal. *Publishers Weekly*, *251(35)*, 11.

Wray, R. (2004). Business: Media Business: MPs back free access to research results. The Guardian Jul 20, 2004 p.17.

Richardson, M. (2005). Open access and institutional repositories: an evidence-based approach. *Serials*, *18*, 98–103.

Rowland, F., Swan, A., Needham, P., Probets, S., Muir, A., Oppenheim, C., O'Brien, A., & Hardy, R. (2004). Delivery, management and access model for e-prints and open access journals. *Serials Review*, *30(4)*, 298-303.

Rowlands, I., & Nicholas, D. (2005). New journal publishing models: an international survey of senior authors. http://www.ucl.ac.uk/ciber/pa_stm_final_report.pdf (accessed 14 January 2006)

Rowlands, I., Nicholas, D., & Huntington, P. (2004). Scholarly communication in the digital environment: what do authors want? *Learned Publishing*, *17*, 261-273.

Rowlands, I., & Nicholas, D. (2005). Scholarly communication in the digital environment : the 2005 survey of journal author behaviour and attitudes. *Aslib Proceedings: New Information Perspectives*, *57*, 481–497.

Royal Society. (2005). Royal Society position statement on 'open access'. http://www.royalsoc.ac.uk/page.asp?id=3882 (accessed 25 November 2005)

Royal Society. (2006). Royal Society launches trial of new 'open access' journal service. http://www.royalsoc.ac.uk/news.asp?id=4838 (accessed 29 September 2006)

Russell, J. (2005). Opening access to uk doctoral theses: the ethos project. *Serials*, *18*, 230.

Sainsbury, D. (2004). Open access is not only science publishing model. Financial Times Nov 10, 2004 p.16.

Schroter, S., Tite, L., & Smith, R. (2005). Perceptions of open access publishing: interviews with journal authors. *BMJ*, *330*, 756–759. doi:10.1136/bmj.38359.695220.82

Shah, S. (2005). Reed Elsevier facing fresh threat to science journal profits. The Independent Jan 6, 2005 p.44.

Shah, S. (2004). Anger as ministers block science publishing shake-up. The Independent Nov 8, 2004 p.38.

SHERPA. (2006). Distribution of UK Higher Education Repositories. http://www.sherpa.ac.uk/documents/rep_distrib.html (accessed 28 September 2006)

Siegel, V. (2004). Open-access publishing. *The Lancet*, *364(9428)*, 25-25.

Simpson, P., & Hey, J. (2006). Repositories for research: southampton's evolving role in the knowledge cycle. *Program*, *40*, 224–231.

SQW. (2004). Costs and business models in scientific research publishing. http://www.wellcome.ac.uk/en/images/costs_business_7955.pdf (accessed 30 November 2005)

Stern, D. (2005). Open access or differential pricing for journals: the road best traveled? *Online*, *29(2)*, 30.

Suber, P. (2003). Scholarly communication – removing barriers to research: an introduction to open access for librarians. *College & Research Libraries News, 64(2),* http://ala.org/ala/acrl/acrlpubs/crlnews/backissues2003/february1/removingbarriers.htm (accessed 2 October 2006)

Suber, P. (2005). Open access overview. http://www.earlham.edu/~peters/fos/overview.htm (accessed 13 November 2005)

Suber, P. (2005). Open access, impact, and demand. *BMJ*, *330*, 1097–1098. doi: 10.1136/bmj.330.7500.1097

Swan, A., & Brown, S. (2004). Authors and open access publishing. *Learned Publishing*, *17*, 219-224.

Swan, A., & Brown, S. (2004). JISC/OSI JOURNAL AUTHORS SURVEY Report. http://cogprints.org/4125/01/JISCOAreport1.pdf (accessed 20 January 2006).

Swan, A., & Brown, S. (2005). Open access self-archiving: an author study. Technical Report, Joint Information Systems Committee (JISC), UK FE and HE funding councils. http://cogprints.org/4385/

Swan, A., Needham, P., Steve, P., Muir, A., Charles, O., O'Brien, A., Hardy, R., Rowland, F., & Brown, S. (2005). Developing a model for e-prints and open access journal content in uk further and higher education. *Learned Publishing*, *18*, 25-40.

Tenopir, C. (2004). Open access alternatives. *Library Journal*, *129(12)*, 33.

Tenopir, C., King, D., & Boyce, P. (2003). Patterns of journal use by scientists through three evolutionary phases. *D-Lib Magazine*, *9 (5)*,

Tenopir, C., & King, D.W. (2002). Reading behaviour and electronic journals. *Learned Publishing*, *15*, 259–265.

Berners-Lee, T., Roure, D.D., Harnad, S., Shadbolt, N., Law, D., Murray-Rust, P., Oppenheim, C., & Wilks, Y. (2005). Open Letter to Research Councils UK: Rebuttal of ALPSP Critique. http://openaccess.eprints.org/index.php?/archives/18-guid.html (accessed 29 November 2005)

Turid, H., Tomas, G., & Bo-Christer, B. (2004). The open access scientific journal: an empirical study. *Learned Publishing*, *17*, 199-209.

Ulrich, P. (2004). Interactive journal concept for improved scientific publishing and quality assurance. *Learned Publishing*, *17*, 105-113.

Spedding, V. (2005). The infrastructure is there: time to populate. http://www.researchinformation.info/special2005selfarchiving.html (accessed 29 November 2005)

Velterop, J. (2003). Should scholarly societies embrace open access (or is it the kiss of death)? *Learned Publishing*, *16*, 167-169.

Velterop, J. (2004). Open access: science publishing as science publishing should be. *Serials Review*, *30(4)*, 308-309.

Vine, R. (2006). Google scholar. *Journal of the Medical Library Association*, *94*, 97–99. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1324783

Walport, M. (2004). The paperless revolution in knowledge. Financial Times. Apr 30, 2004. p.14.

Waltham, M. (2005). Open access – the impact of legislative developments. *Learned Publishing*, *18*, 101–114.

Ware, M. (2004a). Universities' own electronic repositories yet to impact on open access. *Nature webdebates,* http://www.nature.com/nature/focus/accessdebate/4.html (accessed 27 September 2006)

Ware, M. (2004b). Pathfinder research on web-based repositories. http://www.palsgroup.org.uk/ (accessed 29 September 2006)

Ware, M. (2006). Open archives and their impact on journal cancellations. *Learned Publishing*, *19*, 226–229.

Wellcome Trust. (2005). Wellcome Trust position statement in support of open and unrestricted access to published research. http://www.wellcome.ac.uk/doc_WTD002766.html (accessed 29 November 2005).

Willinsky, J. (2003). The nine flavours of open access publishing. *Postgraduate Journal of Medicine*, *49*, 263–267.

Wray, R. (2005). Wellcome boost for open access. The Guardian *Media Business*, 15 December 2005 p.27.

Wray, R. (2005). Springer hires open access pioneer. The Guardian, Aug 17, 2005, p.17.

Wray, R. (2005). Media business: Open access moves a step closer. The Guardian, Feb 2, 2005, p.23.

Wray, R. (2004). Reed allows academics free web access: 200,000 articles may be available on the net but competitors accuse publisher of making token effort. The Guardian Jun 3, 2004 p.23.

Wray, R. (2004). Open access jeopardises academic publishers, Reed chief warns. The Guardian, Jun 30, 2004, p.16.

Wray, R. (2005). Keystroke strategy for open access: Southampton professor will unveil new plan that could revolutionise how academics and the public view research. The Guardian, Mar 1, 2005 p.20.

Wray, R. (2004). Analysis: Confused decision on science publishing. The Guardian Nov 9, 2004, p.16.

Wren, J.D. (2005). Open access and openly accessible: a study of scientific publications shared via the internet. *BMJ*, *330*, 1128–1132. doi: 10.1136/bmj.38422.611736.E0

Wysocki, B.J. (2004). Academic libraries back open access publishing system. Financial Times, May 29, 2004, p.12.

Appendix 1

Web survey

Invitation email

Subject: Survey for MSc dissertation

Dear <Name>

I am currently studying for an MSc in Publishing at Napier University. As part of my dissertation investigating the way in which researchers and end-users find and use journal articles, I am conducting a survey of academic researchers.

I would be very grateful if you could fill in the survey, which can be found at

http://www.tardis.ed.ac.uk/~rhi/Survey1.php

The survey is two pages long and should take no more than ten minutes.

If you have any questions, please email me at this address (04008503@nap.ac.uk). There will be a space for your comments at the end of the survey. If you would like to view a copy of my dissertation when it is complete, please fill in your email address in the box provided in the survey.

Many thanks for your help.

Best wishes

Rhiannon Miller

MSc Publishing, Napier University

04008503@nap.ac.uk

Survey text

MSc Publishing Survey

Thank you for agreeing to take part in this survey. The survey is two pages long, and should take no more than ten minutes. Your email address has not been recorded, but you can if you wish provide it where indicated at the end of the questionnaire in order to receive a copy of the dissertation when it is completed.

The data gathered in this survey will be used as part of a dissertation to be submitted for an MSc in Publishing at Napier University. The dissertation concerns the way people source literature related to their field of study. The results of this survey will be presented in aggregate form, and no personally identifying information will be included.

Rhiannon Miller
MSc Publishing student, Napier University

Start the Survey

Napier University MSc dissertation questionnaire

1. What institution do you belong to? (If you are affiliated to more than one, please name your primary institution.)

2. What sort of institution is this?

- Academic
- Commercial
- Crossover/Partnership
- Government
- Other
- Don't know

3. What field do you work in?

Dropdown menu options:
Biological sciences
Medicine
Chemistry
Physics/Astronomy
Social science
Mathematics
Psychology
Computing/Technology
Languages
Arts
Humanities
Law
Media
Other

Please state if Other:

4. What is your role?

Dropdown menu options:
Postgraduate student
Postdoctoral student
Research assistant
Research Fellow
Lecturer
Professor
Junior partner
Senior partner
Consultant
Other

Please state if Other:

5. How many years have you been working in this field?

- ○ Less than one
- ○ 1-5
- ○ 6-10
- ○ 11-15
- ○ 16-20
- ○ 21-25
- ○ More than 25

6. How many papers have you used (cited or read as part of your research) in the last year?

- ○ None
- ○ 1-10
- ○ 11-50
- ○ 51-100
- ○ More than 100

7. Where do you search for articles online (please tick all that apply)?

- ☐ Google
- ☐ Google Scholar
- ☐ Google Books
- ☐ Medline
- ☐ Emerald
- ☐ Synergy
- ☐ RePEc
- ☐ Edina
- ☐ Proquest
- ☐ FAME
- ☐ CiteSeer
- ☐ CiteBase
- ☐ Ingenta
- ☐ ScienceDirect
- ☐ FindArticles (LookSmart)
- ☐ Cogprints
- ☐ Pubmed Central
- ☐ Biomed Central
- ☐ Scirus
- ☐ ADS
- ☐ CSA or CSA Illumina

- ☐ EBSCO
- ☐ LISA
- ☐ Metapress
- ☐ Ovid
- ☐ Embase
- ☐ SAGE
- ☐ SpringerLink
- ☐ WilsonWeb
- ☐ OAIster
- ☐ Yahoo
- ☐ MSN
- ☐ The website of a specific journal
- ☐ My institution or library internal search facility
- ☐ Other (Please list any other sources you use)

8. Which **one** of these do you use the most?

Dropdown menu includes:
Google
Google Scholar
Google Books
Medline
Emerald
Synergy
RePEc
Edina
Proquest
FAME
CiteSeer
CiteBase
Ingenta
ScienceDirect
FindArticles (LookSmart)
Cogprints
Pubmed Central
Biomed Central
Scirus
ADS
CSA or CSA Illumina
EBSCO
LISA
Metapress
Ovid
Embase

SAGE
SpringerLink
WilsonWeb
OAIster
Yahoo
MSN
The website of a specific journal
My institution or library internal search facility
Other

9. What formats do you prefer for papers you access online? (Please tick all that apply.)

- ☐ HTML
- ☐ XML
- ☐ PDF
- ☐ Word
- ☐ RTF
- ☐ Plain text
- ☐ Postscript
- ☐ Other (Please state) ☐

10. Of these, which one do you prefer the most?

- ○ HTML
- ○ XML
- ○ PDF
- ○ Word
- ○ RTF
- ○ Plain text
- ○ Postscript
- ○ Other (stated above)

One form of open access is author self-archiving, where an article is published on the author's website or placed in a publicly-available database known as a 'repository' in addition to publication in a journal.

11. Did you know about self-archiving before reading this?

- ○ I knew a lot about this
- ○ I knew a little about this
- ○ I did not know about this

12. Please select whether you think the following statements are true or false (if you do not know, please guess):

	True	False
12a. You have to pay to place your article in a repository	○	○
12b. Repositories are only for pre-prints, not for the final article	○	○

	True	False
12c. Repositories are permanent records of research	○	○
12d. Repositories are no better than a personal website	○	○
12e. Publication in repositories undermines journal revenue	○	○
12f. Articles from repositories are as good as articles directly from journals	○	○
12g. Articles from official journal websites are more trustworthy than articles from personal websites	○	○

13. Does your institution or company have an article repository?

○ My institution has a repository for some or all subjects and requires me to place my articles in it

○ My institution has a repository for some or all subjects but does not require me to use it

○ My institution does not have a repository

○ I don't know

14. Have you used articles that you found in an institutional repository (not necessarily your own)?

○ Yes

○ No

15. Do you know of and have you used any subject-based repositories (not limited to any one institution) in your field? (Please tick all that apply.)

☐ I know of at least one subject-based repository in my subject

☐ I have placed articles in at least one subject-based repository in my subject

☐ I have used articles found in at least one subject-based repository in my subject 16. Have you ever done any of the following? (Please tick all that apply.)

☐ posted a copy of the final PDF of an article on my personal website

☐ used an article from an author's website because I didn't have access to the journal it was published in

☐ used an article from an author's website which hadn't been published in a journal

☐ used an article from an author's website because it was less hassle than finding the journal version (even though I had access to it)

☐ used an article from a repository because I didn't have access to the journal it was published in

☐ used an article from a repository which hadn't been published in a journal

☐ used an article from a repository because it was less hassle than finding the journal version (even though I had access to it)

☐ used both the self-archived version and the journal version of a paper because they were different

17. Please indicate how much you agree or disagree with the following statements:

	Disagree strongly	Disagree somewhat	Neutral	Agree somewhat	Agree strongly
17a. I have often found articles relevant to my work in repositories	○	○	○	○	○
17b. I would prefer to cite an article I'd found in a journal rather than one I'd found in a repository or on a website	○	○	○	○	○
17c. I have more confidence in the journal PDF than in the author's own version	○	○	○	○	○
17d. More published articles should be made available through repositories	○	○	○	○	○
17e. It makes no difference to me where I find an article as long as it has been peer-reviewed	○	○	○	○	○
17f. I am happy to wait a year for articles to become available in repositories after they have been published	○	○	○	○	○
17g. It is difficult to tell the difference between pre-print (unrefereed) and peer-reviewed papers on repositories	○	○	○	○	○
17h. Authors should be allowed to publish the PDF of their article online	○	○	○	○	○
17i. I don't pay much attention to the publishers' terms and conditions when I publish a paper	○	○	○	○	○
17j. Readers will find articles through search engines, no matter where they are published	○	○	○	○	○

18. If you have ever published a paper in a journal, what year was your most recent article published in?

○ 2006
○ 2005
○ 2004
○ 2003
○ 2002
○ Earlier
○ I have never been a named author on a published paper

Continue

Napier University MSc dissertation questionnaire: Page 2

19. How many papers have you authored or co-authored in the last year?

- ○ None
- ○ 1
- ○ 2
- ○ 3-5
- ○ 6-10
- ○ More than 10

20. What journal was your last paper published in?

[ ]

21. Were you responsible for submitting the paper to the journal?

- ○ Yes
- ○ No

22. Who owns the copyright?

- ○ I and/or my co-authors own the copyright to my article
- ○ My publisher owns the copyright to my article
- ○ Someone else (e.g. my employer) owns the copyright to my article
- ○ I don't know who owns the copyright to my article

23. According to my agreement with the publisher, I am permitted to do the following with my most recently published paper (please tick all that apply):

- ☐ post a copy of the final PDF on my personal or institutional (publicly-accessible) website
- ☐ post a copy of my own version on my personal or institutional (publicly-accessible) website

- ☐ upload a copy of the final PDF on an institutional or subject-based repository
- ☐ upload a copy of my own version on an institutional or subject-based repository

- ☐ I have to wait until a certain amount of time has passed before I can do (some of) these things
- ☐ I (and my co-authors) didn't sign any publishing agreement
- ☐ I don't know what I am permitted to do with my paper

24. For my most recently published paper, I or my co-authors have done the following (or plan to when the embargo expires): (Please tick all that apply)

- ☐ post a copy of the final PDF on a personal or institutional website
- ☐ post a copy of my own version on a personal or institutional website
- ☐ upload a copy of the final PDF to an institutional or subject-based repository

☐ upload a copy of my own version to an institutional or subject-based repository

☐ I plan to do this when the embargo expires

☐ I am doing something else:

25. If you haven't done, or don't plan to do, any of these things, why not? (Please tick all that apply)

☐ Don't know how/don't have anywhere to upload it
☐ Don't see the need/not required to
☐ Costs money
☐ Don't have the time
☐ Worried that people might steal/plagiarise my work
☐ Worried it might make my paper look lower quality
☐ Don't think my paper will stay available long-term through these means (so no point bothering)
☐ My agreement with my publisher doesn't allow any of these things
☐ I disagree with the idea on principle
☐ My publisher or my institution does it for me, so I don't need to bother
☐ Other:

26. If you have done or plan to do any of these things, why? (Please tick all that apply)

☐ Required to by funding agency
☐ Want to improve my citations
☐ Want to make my work as accessible as possible
☐ Want to encourage other authors to make their work as accessible as possible
☐ Required to by law
☐ Required to by my institution
☐ Want to protect my copyright by asserting ownership
☐ It's no trouble so I might as well
☐ My publisher or my institution does it for me, so I don't have any choice
☐ Other:

27. For the last paper you published that you also self-archived:

○ The self-archived version is identical to the journal version

The self-archived version is different:

○ It has a different layout but identical text
○ It does not contain minor corrections or updates (e.g. spelling/grammar corrections) that **are** in the journal version

It does not contain major corrections or updates (e.g. corrections suggested by reviewers) that **are** in the journal version

It contains minor corrections or updates (e.g. spelling/grammar corrections) that **are not** in the journal version

It contains major corrections or updates (e.g. corrections suggested by reviewers) that **are not** in the journal version

I have never self-archived any of my papers

Please make any comments you wish about the difference between the available versions of your paper:

28. What format(s) did you make it available in? (Please tick all that apply)

HTML

XML

PDF

Word

RTF

Plain text

Postscript

Other (Please state)

29. If you have any comments you would like to add about this survey or the issues it is concerned with, or want to expand on any of your answers, please do so below. Please write as much or as little as you like.

30. If you would like to see a copy of the report when it is completed, please enter your email address. Your address will be used for no other purpose, and will not be sold to or shared with any other organisation.

Submit

Appendix 2

Analysis of institutional repositories

Introduction

I sent the invitation emails to researchers at eight universities, four with and four without institutional repositories. I also sent the email to a small mailing list which mainly consisted of NHS doctors in South Wales. In addition to responses from these institutions, several people affiliated to other institutions also answered the questionnaire.

Although I chose the institutions with the aim of getting a good balance between those with and without repositories, most of the responses came from institutions with repositories. Some reasons for this are explored in the main text. It is worth noting that the number of institutions with repositories is increasing rapidly, with one of the universities implementing a repository during the time that this project was under way.

Survey of institutional repositories

Methods

Fpr each of the institutions represented in the survey responses, I gathered information about the existence of a repository by three means:

- I searched by the name of the institution on OpenDOAR, the database of open-access repositories run by SHERPA (http://www.opendoar.org/),ROAR, the of open-access repositories run by the University of Southampton (http://archives.eprints.org/), and ROARMAP, ROAR's database of repository self-archiving policies (http://www.eprints.org/openaccess/policysignup/).

- I browsed and searched the institutions' websites, looking for an 'archive of research' or 'list of publications' that might link to a repository.

- I searched on Google for the name of the institution together with the word 'repository' or 'archive'.

If none of these means turned up a repository or news of one, I assumed that that institution did not have one.

I then gathered as much information about each repository as I could, from the repository websites themselves and from their entries in ROAR and OpenDOAR. Where contact details were given for the

repository manager, I emailed them with a short survey about the repository. I sent 16 emails to repository managers, and received 10 replies relating to nine institutions.

The text of the email is given in the box below.

Hi there

I am studying for a Masters degree in Publishing at Napier University, and I am currently working on a dissertation on self-archiving and repository use. In particular, I'm looking at whether researchers are aware of and use their institutional repository. It would be enormously helpful if you could answer the following questions about the repository at your institution.

1. When was the repository established?

2. Is the repository intended to cover all subjects (faculties, departments or research fields) in the institution? If not, please could you list the subjects it does cover.

3. Is it mandatory for members of the university to deposit their published articles in the repository? (If it's only mandatory for some of the subjects covered by the repository, please could you list the subjects for which it is required.)

4. Does the repository accept pre-prints, post-prints, or both?

5. What file formats are acceptable, and what is the preferred format?

6. Is there an upload service to deposit papers on behalf of researchers?

7. Has the existence of the repository been publicised to researchers in the university at all? If so, how?

8. Please feel free to add any other information about the repository that you feel would be relevant.

Many thanks for answering these questions. Please let me know if you'd like to see a copy of the

Results

Table 6 shows the repository status for each institution, together with the URL for the repository where it exists. The original institutions to which the main questionnaire email was sent are highlighted in blue.

Institution	Respondents	Repository	Location
Aberdeen	29	Yes	http://auraserv.abdn.ac.uk:9080/aura/
ACM	1	No	
Bournemouth	1	No	
Bristol	1	Yes	http://rose.bris.ac.uk/dspace/
Institute of Cancer Research	2	Coming soon	http://www.sherpa-leap.ac.uk/
Cardiff	53	Yes	http://eprints.cf.ac.uk/
Durham	1	Yes	http://eprints.dur.ac.uk/
Edinburgh	1	Yes	http://www.era.lib.ed.ac.uk/index.jsp
Eurecom	1	No	
Glasgow	1	Yes	http://www.lib.gla.ac.uk/enlighten/
John Innes Centre*	1	No	
Keele	13	No	
Lincoln	22	Yes	http://eprints.lincoln.ac.uk/
Macerata	1	No	
Middlesex	1	Yes	http://eprints.mdx.ac.uk/
MRC*	3	No	
NASA	1	Yes	http://trs-new.jpl.nasa.gov/dspace/
Newcastle	1	Pilot	http://rogue.ncl.ac.uk/
NICTA	1	Informal*	http://www.ertos.nicta.com.au/publications/
NHS	9	No	
Oxford	1	Yes	http://eprints.ouls.ox.ac.uk/
Reading	33	No	
Robert Gordon	1	No	
Roehampton	13	Coming soon	http://rrp.roehampton.ac.uk/
Rutherford Appleton Laboratory	2	Yes	http://epubs.cclrc.ac.uk/
Southampton	76	Yes	http://eprints.soton.ac.uk/, http://eprints.ecs.soton.ac.uk/
Staffordshire	1	No	
Surrey	5	Yes	http://epubs.surrey.ac.uk/
Thessaloniki	1	Yes	http://cds.lib.auth.gr/
UCL	41	Yes	http://eprints.ucl.ac.uk/
UEA	63	Coming soon	

*Does not have an institutional repository of its own but will mandate self-archiving in another repository from October 2006.
†NICTA has a webpage with a list of published papers with full-text PDFs where possible (mostly technical reports and conference papers) rather than a formal (OAI-compliant) repository.

Table 6: Repository status of institutions

Roehampton's repository went online during the course of this project. As it has not apparently been officially launched, and as of 28 September there were only 13 articles archived in it, I counted Roehampton as not having a repository for the purpose of analysing the data obtained in this survey, In addition, the University of East Anglia and the Institute of Cancer Research are actively developing repositories and have started publicising them. There are therefore only 11 institutions out of the 34 in this study that do not have either a current repository or plans to develop one. Moreover, three of the UK Research Councils are represented in this table: one respondent gave their affiliation as the John Innes Centre, which is run by the Biotechnology and Biological Sciences Research Council (BBSRC);

two were from the Rutherford Appleton Laboratory, run by the Council for the Central Laboratory of the Research Councils (CCLRC); and three were from the Medical Research Council (MRC). Of these, the Rutherford Appleton Centre, through the CCLRC, has a repository and is listed as such in the table; the BBSRC and the MRC do not but, in accordance with the other Research Councils, are introducing in October 2006 a policy of mandatory self-archiving (though this will not apply to the current research of respondents to this survey as the mandate has not yet come into effect). The BBSRC does not specify a repository, but the MRC will require that research output be placed in PubMed Central in the first instance, and, when it is available, PubMed Central UK.

Two of the institutions had more than one repository relevant to the respondents in this survey[12]. The University of Southampton has a separate repository for the computing science department in addition to its main repository. The University of Glasgow has two separate repositories within their overall repository service, Enlighten. One repository is for pre-print (non-peer-reviewed) papers and there is a separate repository for peer-reviewed, final publication versions. The ePrints server, which is for final published papers, was launched in 2004 and holds over seven times the number of articles in the DSpace server, which holds the pre-prints and, though live, has not yet been officially launched.

Table 7 shows the characteristics of the institutional repositories listed in Table 6.

Institution	Established	Subjects	Number of items*	Mandatory	Preferred format	Accepts
Aberdeen	Oct 2005	All	52	no	PDF	Postprints
Bristol	Jun 2005	All	728	no (but under consideration)	PDF	Both
ICR	Jan 2006†	All		no	PDF only	Postprints
Cardiff	Feb 2006	All	62	no	PDF	Both
Durham	Dec 2003	All	166	no		Both
Edinburgh	Jan 2005	All	1017	no	PDF	Both
Glasgow	Feb 2004	All	3139	no	PDF, Word	Both
Lincoln	Sep 2004	Computing	80	no	PDF	Both
Middlesex	Unknown	Unknown	29	no	unknown	unknown
NASA	unknown	All	20877	no	unknown	Both
Newcastle	Nov 2004†	All	286	no		Both
Oxford	Jun 2004	All	546	no	PDF, HTML, Word	Both
Rutherford Appleton	May 2004	All	23901	no (but soon will be)	Any	Both
Southampton	Feb 2003	All	31089‡	no (but soon will be)	Any	Both
Surrey	Dec 2004†	All	318	No (but under consideration)	PDF	Both
Thessaloniki	Apr 2003	All	54748		PDF	Unknown
UCL	Jul 2004	All	1853	no		Both
UEA	2007§	All		no	PDF	Postprints

* Data from ROAR and individual repository websites; valid as of 29 September 2006.
† Still in pilot phase.
‡ Aggregate of two repositories: ePrints Soton (all departments, 20513 items) and the University of

[12] Some institutions, such as Oxford University and NASA, also had other repositories, but these were not applicable to the respondents from that institution and I did not consider these in this survey.

Southampton: Department of Electronics and Computer Science departmental repository (10546 items).
§ Projected.

Table 7: Properties of institutional repositories

The large variation in repository size, as measured by the number of items listed in each, is likely to be related to each institution's size and the length of time the repository has been operational, but could also reflect the number of different types of documents (journal articles, theses, conference papers, technical reports, etc.) allowed or encouraged by the institution to be self-archived. In addition, some repositories allow and encourage the retroactive self-archiving of researchers' past output, whereas others concentrate on new publications only. I did not consider the relative sizes of repositories in my analysis of the effect of the presence of a repository on respondents' attitudes; the majority of respondents came from institutions with small repositories.

Of those for which information was available, most repositories accepted both preprints (non-peer-reviewed works) and postprints. All accepted PDF files and most accepted other files as well, though PDF was the preferred format for those that accepted other formats, and some (e.g. Glasgow) converted files received in other formats to PDF.

Almost all institutions had repositories that were intended to receive articles from all departments, although many had only been used by researchers from a few subject areas, particularly where the repository had only recently been set up or was in a pilot phase. The main exception, Lincoln University, currently has a repository only for the Department of Computing and Informatics, but is planning an institution-wide repository which will absorb the current archive.

No institution in this survey currently had a policy of requiring all its members to deposit all their research output in its repository. The University of Southampton's departmental repository for the Department of Electronics and Computer Science does have such a mandate, and the University is in the process of implementing one for the institution-wide repository as well, and this will be active very shortly. Along with some other institutions, the University of Southampton is also conducting its Research Assessment Exercise (RAE) review through the repository, thus obliging researchers to archive their best work in the repository in order for it to be considered in the review (IRRA Project, 2006 Simpson and Hey, 2006). A mandatory policy is also being considered for Bristol University's repository.

Most institutions publicised their repositories by various means, with mentions in meetings and articles in newsletters being the most popular forms of publicity (Table 8), as well as information and advocacy on the repository website, often with a list of the advantages in terms of publicity and citations that would accrue from self-archiving research output. Most also had a service to assist researchers with upload of their papers, to encourage self-archiving by reducing the work involved, particularly where retroactive archiving was being promoted.

Institution	Upload Service	Publicised
Aberdeen	yes	leaflets, advocacy
Bristol	no	newsletters, emails, meetings
ICR	no	newsletters, meetings
Cardiff	yes	advocacy
Durham	no	
Edinburgh	yes	
Glasgow	yes	newsletters, events, leaflets, presentations
Lincoln	no	word of mouth, presentations
Middlesex	unknown	
NASA	Unknown	
Newcastle	no	
Oxford	yes	
Rutherford	yes	Leaflets, presentations, freebies
Southampton	no	RAE, meetings, presentations, training
Surrey	yes	Surveys, meetings
Thessaloniki		
UCL	yes	
UEA	yes	Meetings

Table 8: Publicity of repositories